Doctoral Thesis

Improvement of Air Quality Forecasts based on Aerosol Data Assimilation with Satellite and Ground Observations

Ganghan Kim

Department of Urban and Environmental Engineering
(Environmental Science and Engineering)

Ulsan National Institute of Science and Technology

2021

Improvement of Air Quality Forecasts based on Aerosol Data Assimilation with Satellite and Ground Observations

Ganghan Kim

Department of Urban and Environmental Engineering

(Environmental Science and Engineering)

Ulsan National Institute of Science and Technology

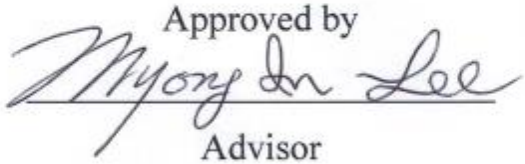
Improvement of Air Quality Forecasts based on Aerosol Data Assimilation with Satellite and Ground Observations

A dissertation submitted to UNIST
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Ganghan Kim

06/15/2021 of submission

Approved by



Advisor

Myong-In Lee

Improvement of Air Quality Forecasts based on Aerosol Data Assimilation with Satellite and Ground Observations

Ganghan Kim

This certifies that the dissertation of Ganghan Kim is approved.

June/15th/2021

Signature

Advisor: Myong-In Lee

Signature

typed name: Chang-Keun Song

Signature

typed name: Sung Deuk Choi

Signature

typed name: Cheol-Hee Kim

Signature

typed name: Rokjin Park

Abstract

This study develops a 3DVAR data assimilation and forecasting system that simultaneously assimilates aerosol optical depth (AOD) from Geostationary Ocean Color Imager (GOCI) and Moderate Resolution Imaging Spectroradiometer (MODIS) satellites and ground-based PM_{10} and $PM_{2.5}$ observations using the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem). The simulation domain covers Northeast Asia, and the assimilation and forecast skill is evaluated for two periods, one is Korea-US Air Quality (KORUS-AQ) intensive observing period and the other is April of 2017. In evaluating the data assimilation performance, the assimilated surface PM concentrations exhibit higher consistency with the observed data by showing increased correlations for PM_{10} and $PM_{2.5}$ from the no assimilation run. The data assimilation also shows beneficial impacts on the air quality forecasts over South Korea until 24 hours from the initialized states. A couple of deficiencies are also found in the data assimilation and forecast system. They show pronounced seasonal dependence in the forecast skill, suggesting an important role of the seasonal changes in regional atmospheric circulation patterns. The forecast accuracy becomes improved than the background model statement, which is most of the skill improvement comes from the surface PM assimilation, showing a practically useful level of skill until 12 hours from the initial state. The seasonal forecast becomes lower in spring and winter when the ambient aerosol concentration is higher than in other seasons. The forecasts also show much degradation as the forecast lead time increases due to systematic model biases. When a simple statistical bias correction is applied to the forecast outputs by adjusting the mean and variance of the forecast outputs to those in the observed distribution, the forecast maintains the skill at a practically useful level for more than a day. For a categorical forecast, the skill score has increased by more than 10 % on average and 37 % for high-concentration events, respectively. Additionally, the analysis of the impact of the data assimilation component for the 3DVAR has been done. Each is a data assimilation method compared to EnKF, the observing system experiments (OSE), and the adaptation of Four-Dimensional Data Assimilation (FDDA) for the constraining meteorology. EnKF method showed comparable or less data assimilation skills because of the spread problem, OSE showed the partial data assimilation effect to the different types of variables and the slight improvement of the forecast accuracy. The FDDA simulation showed that the constraining meteorology can affect the atmospheric aerosol directly by the wind modification, and in this study, the PM result has been improved because of the FDDA option.

Contents

List of Figures

List of Tables

Chapter 1

1. Introduction

1.1. Background and Motivations

The atmospheric aerosol research has been conducted for dozens of years which is directly related to human health. The concentration of atmospheric aerosol affects human health and fatality (Pope Iii et al. 2002). Previous research described that the fatality related to the atmospheric aerosol concentration, when the concentration has been elevated to 10 $\mu g/m^3$, the fatality risk related to all-cause, cardiopulmonary, and lung cancer becomes increased to 4%, 6%, and 8%. Other studies analyzed the concentrations of particulate matters and respiratory system, which are by Tie et al. (2009), increasing aerosol extinction coefficients throughout the past years' significant increment of lung cancer mortality 10 to 70 deaths per 100,000 people from 1960 to 2002 when atmospheric aerosol increased from 170 to almost 400 M/m. To prevent the health affect from atmospheric aerosols, not only the reduction of aerosol emission is important, but also monitoring the transport of atmospheric aerosols are necessary.

The air quality monitoring is done by several installments over surface stations and from satellites. One of the global aerosol observing networks is Aerosol Robotic Network (AERONET; Holben et al. 1998). The installment is the sun photometer installed over a surface station, and 119 stations are observing atmospheric aerosols for 2020. Most of the countries operating their surface air quality monitoring stations to observe and archive the climatological air quality data. For example, South Korea has a National Ambient Air Quality Monitoring System (NAMIS) operated by the National Institute of Environmental Research (NIER), China has an air quality monitoring network called China National Environmental Monitoring Center (CNEMC). Surface monitoring stations can provide the most accurate pollutant concentrations for the regional area, however, these data have narrow horizontal representativeness because of the point-scale observations. Other types of observation data monitor the air pollutants can observe the wide range. Satellites can observe a broad swath of the region based on their reflected signals for certain wavelengths. There are various kinds of satellites to observe the Earth, and some of the satellites have multiple wavebands to observe not only the meteorological motions and surfaces but also observe atmospheric aerosols. Moderate Imaging Spectroradiometer (MODIS) satellites can observe the atmospheric aerosols as aerosol optical depth (AOD) and can observe the entire Earth atmosphere. MODIS is installed in Aqua and Terra satellites and orbits with polar orbit. For

Northeast Asia, there are several geostationary satellites have been deployed, and continuously monitor the fixed region over time. Geostationary Ocean Color Imager (GOCI), Meteorological Imager (MI), Himawari-8 satellites currently monitoring Korea, China, and Japan. Satellite data is quite useful data, unlike surface station data, it can provide regional representation. However, the satellite cannot observe the single region for a continuous period. Even the geostationary satellite, which can monitor continuously for a certain region, cannot monitor the region for whole-24-hours. These satellites installed passive imager, which only can gather the reflected radiance signals and only can observe in the daytime. Polar orbit satellites can observe in the nighttime, but they only can monitor a certain time for a certain region. Observation data is the most accurate data, but for the continuity perspective, monitoring with only observation might not be enough. Because of that, to support the observation network, numerical modeling is selected as a powerful supporting system. A numerical modeling system contains the various physics and dynamics to simulate the atmospheric movements, which might not as accurate as observations but can simulate the unobservable regions by observation.

There are several numerical models which are capable to simulate atmospheric chemical transportation. Community models such as Weather Research and Forecasting Chemistry coupled model (WRF-Chem; Grell et al. 2005), California Puff model (CALPUFF; Scire et al. 2000), and The American Meteorological Society/Environmental Protection Agency Regulatory Model (AERMOD; Cimorelli et al. 2005) are provided. Each of the models has its unique characteristics and the applicable range is also different. Many global studies have been conducted several experiments using those dispersion models, such as Seangkiatiyuth et al. (2011), who established the AERMOD simulation using multiple pollutant dispersions including NO_x, PM, and SO_2. Mokhtar et al. (2014) conducted a health risk assessment of the surroundings of the coal-fired power plant by AERMOD modeling. For the CALPUFF case, Ghannam and El-Fadel (2013) established the industrial complex simulation with CALPUFF, based on MM5 3D-meteorological data. WRF-Chem is adapted by numerous researchers and studies, such as Lee et al. (2020) designed the experiment to quantify the aerosol activation and nucleation for winter Storms in California. There are also the other kinds of atmospheric transportation models are existing as well, such as GEOS-Chem and Ozone and Related Chemical Tracers (MOZART; Brasseur et al. 1998), but in this study, WRF-Chem takes the majority.

WRF model is one of the most popular community models to simulate three-dimensional meteorology. Coupled with chemistry, the WRF-Chem model can process the massive emission and transport of atmospheric pollutants with various options of chemical schematics. However, similar to the other numerical models, WRF-Chem also consists of several problems, and the most important issue is the model accuracy issue. Current numerical models have their systematic errors because of the imperfect

internal equations about atmospheric physics and dynamics. The operational meteorology and chemistry forecast continues over a day. For example, the forecast over South Korea lasts from 3 days to 1 week (https://www.airkorea.or.kr/web/dustForecast?pMENU_NO=113). These errors from the model system continuously deteriorate the results during the longer forecasting periods. This error can be partially resolved by improving the initial condition of the model, and this feature is directly related to the data assimilation method.

Data assimilation is one of the most sophisticated methods to improve the model result. The basic idea of the data assimilation is "Using all the available information, to determine as accurately as possible the state of the atmospheric flow." (Talagrand 1997). Using all available information is determined for both observation and model data, to produce a more realistic gridded analysis. The adaptation of the data assimilation method has been done first to the atmospheric field to generate the reanalysis and forecast data such as National Centers for Environmental Prediction (NCEP) Final Reanalysis (FNL; Commerce 2000). Currently, the idea of data assimilation is accepting the limitations of both observation and models and merging both data to overcome the limitations and obtain valuable analysis data. Figure 1-1 describes the concept of data assimilation based on the case study of the KORUS-AQ scenario. GOCI and Model show significantly different patterns and data availability, the assimilated analysis in 3rd figure shows merged pattern from both observation and model.

The adaptation of the data assimilation has been considered since the year 2001. Collins et al. (2001) primarily generated the assimilated aerosol from the installment of Advanced Very High-Resolution Radiometers (AVHRR; Stowe et al. 1997) on NOAA 14 satellite with optimal interpolation method. The primary data assimilation showed a significant improvement of output data, and the advanced data assimilation method has been adapted to the operational forecasts. Europe Radar Conference (EURAD) system applied data assimilation of multiple type of observation over Europe region. For East Asia, Liu et al. (2011) has been adapted 3DVAR data assimilation with GOCART aerosol module. This study adapted Community Radiative Transfer Model (CRTM; Weng et al. 2005) to properly assimilate the MODIS AOD. In addition, there are several improving studies such as defining the observational operator equation for PM (Pagowski et al. 2010), using GOCI AOD to assimilate aerosol (Saide et al. 2014), using Fengyun-4A satellite (Xia et al. 2020), and more studies are conducted to generate the better quality of atmospheric air quality even to the current period. Currently, the data assimilation has been adapted by several government ministries and environmental institutes to produce a better forecast of air qualities based on the various kinds of models and data assimilation. The operating systems are described in Table 1-1.

Although previous studies have demonstrated the beneficial impacts of data assimilation to the in-situ

and satellite aerosol observations, most of the studies remained at the test level based on a short testing period. Considering that the data assimilation and forecast performance may depend on regions and testing period, the performance in a specific system should be evaluated based on a rigorous test with a sufficiently long evaluation period. Particularly in East Asia, where the summer and winter monsoon produce distinctive characteristics in regional atmospheric circulation and aerosol transport, the forecast performance is likely to vary according to the season. In addition, natural and anthropogenic emissions show a pronounced seasonality. For example, the regional air quality is manifested by Asian Dust episodes in spring, while coal-fired heating in China generates large amounts of hazardous emissions and degrades air quality in winter. The operational forecast system should demonstrate reliability in all seasons. Also, the aforementioned studies demonstrated well the benefits of using satellite-derived AOD data in improving analysis and, thereby improving air quality forecast skills. For the regions where the long-range transport of dust and chemical pollutants is dominant such as in Northeast Asia including China, Korea, and Japan, the satellite data are ideal due to their wide areal coverage. On the other hand, the conclusions made from the previous studies need to be considered carefully, as they have been made based on various sources of observational data and with different model configurations among the studies. In particular, the individual contributions from the satellite-derived and in-situ observations have not been assessed quantitatively in a single data assimilation system. For example, the study by Liu et al. (2011) and Saide et al. (2014) used the satellite-derived AOD products only, and the study by Li et al. (2013) used the in-situ $PM_{2.5}$ observations only. Moreover, most studies remained to demonstrate the improvement for the specific cases (Liu et al. 2011) or a short testing period of fewer than two weeks (Saide et al. 2014). In this regard, it needs a more extended test for the quantitative assessment of the improvement.

Table 1-1. International atmospheric chemical data assimilation and forecasting systems operated by governments and environmental institutes.

Model (Institute)	Horizontal Resolution	Vertical Levels and Top	Assimilation Method	Assimilating Variable
CHIMERE (INERIS, CNRS)	25 km	8 levels 500 hPa	Kriging	O_3, NO_2, PMs
EURAD-IM (RIU)	15 km	23 levels 10 hPa	3DVAR	NO_2, SO_2, CO, O_3, PMs
LOTOS-EUROS (KNMI, TNO)	25 km	4 levels 3.5 km	EnKF	O_3, NO_2, PMs
MATCH (SMHI)	0.2 °	40 levels 100 hPa	3DVAR	NO_2, SO_2, CO, O_3, PMs
MOCAGE (Météo-Fr, CERFACS)	0.2 °	47 levels 5 hPa	3DVAR	O_3, NO_2, PM_{10}
SILAM (FMI)	0.2 °	46 levels 100 hPa	4DVAR	NO_2, SO_2, CO, O_3, PMs
RAQMS (NOAA)	2 ° (Low-res) 110 km (Hi-res)	26 levels, 60 km (Low-res) 50 levels, 20 km (Hi-res)	3DVAR	AOD, CO, O_3, NO_2

1.2. Research Objectives

In this thesis, each chapter consist of 3DVAR atmospheric aerosol data assimilation testing and analysis as the experiment flows. The study begins with the establishment of the system, verification, analysis, and forecast comparison with KORUS-AQ and long-term analysis, and ends with sensitivity test and analysis of the impact of the data assimilation components.

The study has been conducted from various general questions about the relationship between data assimilation and providing observation-reliable analyses. It starts from searching the value of the aerosol data assimilation. From the procedure, the level of improvement based on 3DVAR data assimilation for the aerosol analysis fields and the forecast has been tested by specific observation period over Northeast Asia and seasonal characteristics for a year. From the 1-year seasonal simulation, differences in data assimilation and forecast quality between seasons also have been searched. From the results, an investigation of the reasons for the difference has been analyzed.

After the general verification, several data assimilation components which affect the actual quality of the data assimilation have been tested. The characteristics of the data assimilation component affecting the quality of the data assimilation have been tested and analyzed. Each of the components is the data assimilation method, the data type of the observations, and the meteorological constraining inside the numerical model.

In chapter 2, the Data and Methods of this study will be described. In Chapter 3, results and analyses will be described. The first subject is the basic 3DVAR data assimilation and forecast system based on WRF-Chem has been described and primary data assimilation skill has been verified. The cases are categorized as KORUS-AQ cases and seasonal differences of 2017 cases. 2017 case is a long-term simulation categorized as season and analyzes the data assimilation and forecast skills for each month. Chapter 4 describes the sensitivity test for various options of data assimilation to determine the optimal configuration for Northeast Asia. In addition, the further plans that describe the improved data assimilation method, Ensemble Kalman Filter (EnKF; Evensen 2004) experiments will be described.

For the KORUS-AQ cases, methods, experiments, analysis, and forecast results have been referenced by Kim et al. (2021)

2. Chapter 2

Data and Method

2.1. Model Configuration

The air quality model used in this study is the WRF-Chem version 3.9.1. WRF-Chem is an online model with coupled interaction between atmospheric chemistry and meteorology. The model considers the direct and indirect effects of atmospheric chemicals and aerosols on atmospheric radiative transfer and cloud microphysical processes. The model uses the Goddard Chemistry Aerosol Radiation and Transport (GOCART; Chin et al. 1999) as an aerosol scheme, including 15 types of dust, organic carbon (OC), black carbon (BC), sulfate, and sea salt. Atmospheric chemistry is formulated by the MOZART. Parameterized physics include the Lin scheme (Lin et al. 1983) for microphysics, the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997) scheme for longwave radiation, the Goddard scheme (Chou and Suarez 1994) for shortwave radiation, the Grell-3D scheme (Grell and Dévényi 2002) for convection, the Yonsei University (YSU; Hong et al. 2006) scheme for atmospheric boundary layer turbulence, and Noah Land-Surface Model (Chen et al. 1996) for land surface physics.

The main aerosol model for this study is the GOCART model. It contains 15 aerosol species categorized as bulk-bin hybrid species. GOCART can simulate black carbon (BC), organic carbon (OC), sulfate, dust, and sea salts. BC and OC are categorized by hydrophobic (BC1, OC1) and hydrophilic (BC2, OC2). Dust and sea salts are divided by their radius. Dust particles are divided by the radius of 0.6~1, 1~1.8, 1.8~3, 3~6, and 6~10 micrometer, and sea salt particles are divided by the radius of 0.1~0.5, 0.5~1.5, 1.5~5, and 5~10 micrometer. The other particulate matters are categorized as p25 and p10, which are unspeciated $PM_{2.5}$ and PM_{10}. GOCART contains its natural dust scheme, which can emit dust from surface clay, sand, and erosion parameters. The natural dust scheme of GOCART is based on the study of Ginoux et al. (2001) The dust emission is drifted by the wind flux of 10 m height, the uplift flux has occurred when the 10 m wind is faster than the threshold wind. The flux F_p of particle size class p is approximated by the equation below.

$$F_p = CSs_p u_{10m}^2 (u_{10m} - u_t)$$

The C is a dimensional factor equal to 1 μg s^2 m^{-5}, S is the source function, u_{10m} is the horizontal wind

speed at 10 m, u_t is the threshold wind speed, and s_p is the fraction of each size class.

The fraction of clay and silt sizes are different for different soil types at each location. These are pre-defined values, in which the s_p values are 0.1 for 0.1-1 μm, 1/3 for each of the classes 1-1.8 μm, 1.8-3 μm, and 3-6 μm, respectively.

The dust source is defined by clay and sands, and the source function S, which is the fraction of alluvium available for wind erosion is defined as the equation below.

$$S = \left(\frac{z_{max} - z_i}{z_{max} - z_{min}}\right)^5$$

z_{max} and z_{min} are the maximum and minimum elevations in the surrounding 10 x 10 topographical region, z_i is a certain elevation for grid i.

The threshold wind is computed from the particle size and the soil moisture. The equation to compute the threshold wind velocity is described below.

$$u_t = A\sqrt{\frac{\rho_p - \rho_a}{\rho_a} g \Phi_p (1.2 + 0.2 \log_{10} w)}$$

A is a dimensionless parameter equal to 6.5, ρ_p and ρ_a are particles and air density, g is the acceleration of gravity, Φ_p is the particle diameter, and w is the surface wetness ranges from 0.001 to 1. The equation only operated when the $w < 0.5$. Otherwise, u_t becomes infinite and no dust flux is processed from the surface wind.

MOZART is the chemical model based on ozone tracers. The model installed in WRF-Chem consists of 81 chemical species including nitrogen compounds (NO, NO_2, NO_3, N_2O_5, N_2O), carbon compounds (CO, HCHO, CH_3OH, C_2H_4, C_3H_6, …), ozone, oxides including sulfur dioxide. SO_2 and sulfate are directly related to secondary aerosol formation.

Every atmospheric chemical becomes removed from the atmosphere by various kinds of aging processes such as dry deposition, chemical reaction, and change to another species. In GOCART, sulfate is processed from SO2, hydrophobic carbons are transferred to the hydrophilic carbons, and dry/wet depositions are applied to derive genuine atmospheric removal of aerosol. Advection is computed by a flux-form semi-Lagrangian method (Lin and Rood 1996). Boundary layer turbulent mixing is treated by a second-order closure scheme (Helfand and Labraga 1988). Moist convection is calculated using the cloud mass flux archived in the GEOS DAS data. Dry deposition includes gravitational settling as a function of particle size and air viscosity and surface deposition as a function of surface type and

meteorological conditions (Wesely and Lesht 1989). Wet deposition accounts for the scavenging of aerosols in convective updrafts and rainout/washout in large-scale precipitation (Balkanski et al. 1993; Giorgi and Chameides 1986). Chemical reactions, including reactions of DMS and SO2 with OH in the air and SO2 with H2O2 in the cloud, are calculated using prescribed oxidant fields from the IMAGES model (Müller and Brasseur 1995).

Figure 2-1 shows the model domain in East Asia (D01) at 27-km horizontal resolution nested in one way by a sub-domain (D02) at 9-km resolution just over the Korean peninsula. This domain has been adapted to the 2017 case simulation. Figure 2-2 shows the model domain similar to Figure 2-1, but for 15-km horizontal resolution and only for domain 1. This domain has been adapted to the KORUS-AQ case simulation. The model is configured with 30 vertical layers with the top layer at 50 hPa for the KORUS-AQ and All-Season case. The initial and lateral boundary conditions (ICs and BCs) for meteorology are specified using the Met Office United Model (UM; Bellouin et al. 2011) analysis produced by the Korea Meteorological Administration. The chemical ICs and BCs are specified by the MOZART-4 (Emmons et al. 2010) global chemistry reanalysis data produced by the Goddard Earth Observation System 5 (GEOS-5) global model coupled with MOZART. The model uses the Konkuk University's Comprehensive Regional Emissions inventory for Atmospheric Transport Experiment (KU-CREATE; Woo et al. 2020) for anthropogenic emission for gases and aerosols for optimum simulation of the 2017 case. For KORUS-AQ and sensitivity test, the Emissions Database for Global Atmospheric Research-Hemispheric Transport of Air Pollution (EDGAR-HTAP; Janssens-Maenhout et al. 2012) has been selected as anthropogenic emission. The model also includes the biological emissions from the Model of Emissions of Gases and Aerosols from Nature (MEGAN; Guenther et al. 2006). For the KORUS-AQ case, the Fire Inventory developed by NCAR (FINN; Wiedinmyer et al. 2011) has been adapted.

UM is the global meteorological data processed by Met Office, UK. UM – Global Data Assimilation and Prediction System (GDAPS) is currently applied to the forecasting system operated by the Korean Meteorological Agency (KMA) and National Institute of Environmental Research (NIER), which are the meteorological and chemical transporting forecasts. UM-GDAPS has a 10 km resolution, 70 vertical levels to 80-km altitude, and data is divided by the 3-hour interval.

EDGAR-HTAP is a global emission inventory generated by the European Commission (EC). It includes nitrogen oxides, sulfuric oxide, methane, NH3, VOC, CO, and carbonaceous aerosols. The adapted version is version 2, the data is based on 2010 global emission research data of basic EDGAR, Model Inter-Comparison Study for Asia (MICS-Asia; Carmichael et al. 2008), Environmental Protection Agency-United States collaborated with Canada (EPA-US/Canada), and Cooperative

Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe-Toegepast Natuurwetenschappelijk Onderzoek (EMEP-TNO; Vestreng et al. 2007). For the current study, the EDGAR-HTAP emission surrounding Northeast Asia. Sample patterns of the aerosol emission from EDGAR-HTAP have been shown in Figure 2-3.

MEGAN is the biochemical model that can generate the biological emission inventory from leaf area index, weather, and atmospheric chemical composition. The emission model has been developed by NCAR and acts as a stand-alone emission generator model for WRF-Chem. The data is differentiated by leaf area index and it is variated by month. The base resolution of MEGAN is about 1 km. MEGAN generally contains the global organic carbon emission.

FINN is the fire emission inventory developed by NCAR. FINN is generated based on satellite observations and land cover information. Fire emission is mostly depending on the actual wildfire events, each of the periods contains each of the unique fire events observed by the global satellite. FINN mostly considers the black carbon emission from fire.

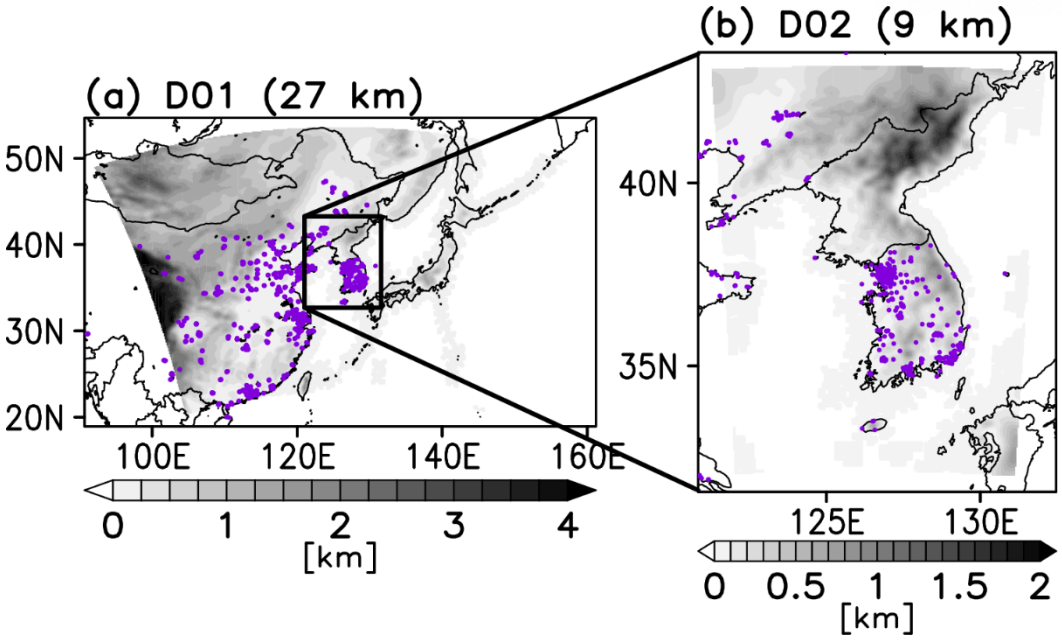


Figure 2-1. Simulation domain for WRF-Chem with surface elevation (shaded, unit: meter). (a) shows domain 1 (D01), and (b) shows domain 2 (D02). Purple dots indicate the locations of surface PM observations.

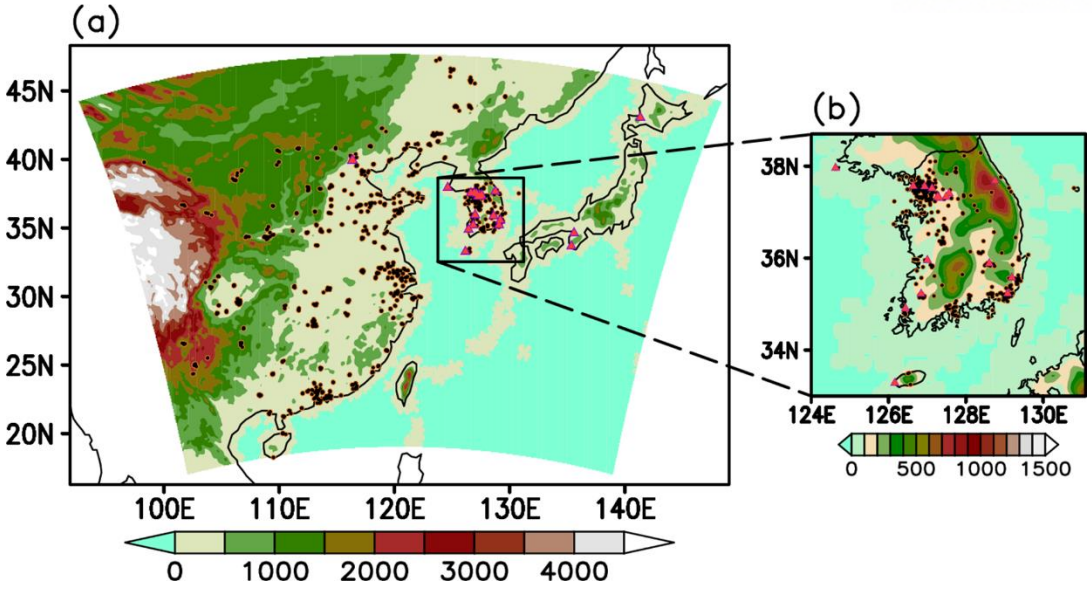


Figure 2-2. (a) The simulation domain in Northeast Asia by WRF-Chem (only shaded region) is shown with surface elevation (unit: meter). The model uses the Lambert Conformal map projection. (b) shows the validation domain over the Korean Peninsula. The orange circles in the figures indicate the ground PM observation sites, and the red triangles the AERONET sites for AOD observations.

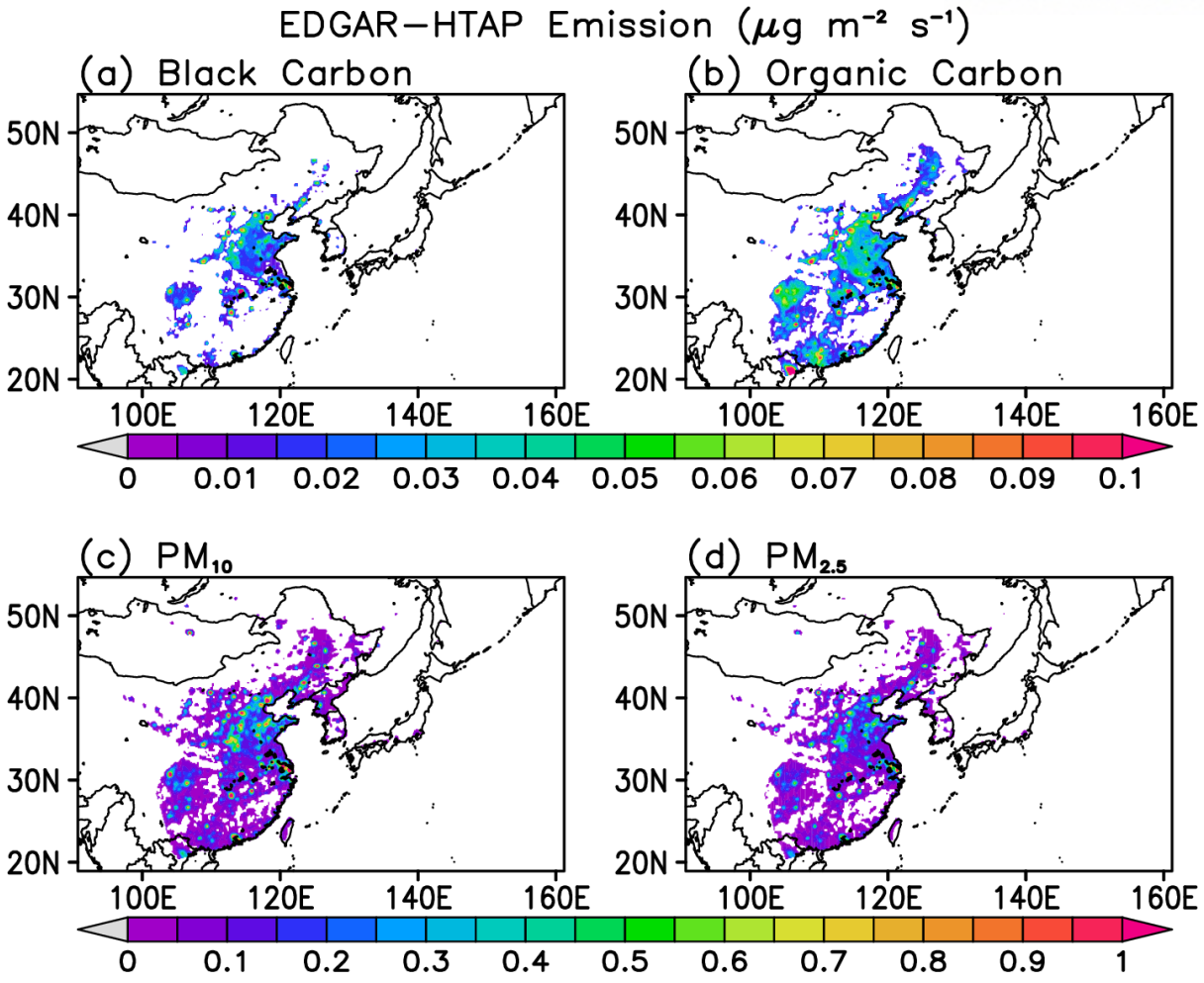


Figure 2-3. The sample EDGAR-HTAP aerosol emissions over the Northeast Asia domain. Each are (a) Black Carbon, (b) Organic Carbon, (c) PM_{10}, and (d) $PM_{2.5}$.

2.2. Data Assimilation Methods

The data assimilation method is the 3D-VAR algorithm implemented by the Grid-point Statistical Interpolation (GSI) version 3.5 (Kleist et al. 2009). Details in the 3D-VAR application to the GOCART aerosol scheme are described in Liu et al. (2011), and only the essential part is described briefly here.

The 3D-VAR algorithm is to find the best estimate of analysis by considering the Euclidian distance from observations at irregularly spaced points and a gridded background field. The analysis vector (x) can be determined by minimizing a scalar cost function J (x) given by

$$\mathrm{J(x)} = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}[H(x) - y]^T R^{-1}[H(x) - y]. \quad (1)$$

In Eq. (1), x_b denotes the model background state, y is the vector of observations, and B and R are the background and observation error covariance matrices, respectively, to represent the relative weighting to analysis. H is an observation operator that converts the model values to observations and interpolates the model values to the observation positions in space. The AOD observation operator is calculated based on the CRTM, and the observation operators of PM_{10} and $PM_{2.5}$ are described by Pagowski et al. (2010).

Finding the smallest cost function J can define the analysis x. However, in the numerical model, the general cost function is unable to compute because of the background error covariance B. It is impossible to compute the inverse matrix of B, which the general dimension of B is about 10^9. To avoid the problem, finding the zero derivatives of cost function has been suggested, and the iteration method has been prepared to prevent compute inverse covariance matrix. For the iteration process, the original cost function becomes transformed into the directional variables.

From equation 1, Define an analysis increment (Δx=) $\boldsymbol{X} = x - x_b$, then equation (1) becomes:

$$\mathrm{J} = \frac{1}{2}\boldsymbol{X}^T B^{-1}\boldsymbol{X} + \frac{1}{2}[H(x_b + \boldsymbol{X}) - y]^T R^{-1}[H(x_b + \boldsymbol{X}) - y]. \quad (2)$$

By assuming the linearity of the observation operator H, equation (2) can be written as:

$$\mathrm{J} = \frac{1}{2}\boldsymbol{X}^T B^{-1}\boldsymbol{X} + \frac{1}{2}\big[H(\boldsymbol{X}) - \big(y - H(x_b)\big)\big]^T R^{-1}\big[H(\boldsymbol{xX}) - \big(y - H(x_b)\big)\big]. \quad (3)$$

Next, define the observation innovation as $\boldsymbol{O} = y - H(x_b)$, equation (3) becomes:

$$\mathrm{J} = \frac{1}{2}\boldsymbol{X}^T B^{-1}\boldsymbol{X} + \frac{1}{2}[H(\boldsymbol{X}) - \boldsymbol{O}]^T R^{-1}[H(\boldsymbol{X}) - \boldsymbol{O}]. \quad (4)$$

To improve convergence, the 3DVAR system preconditions its cost function by defining a new variable of $\boldsymbol{Y} = B^{-1}\boldsymbol{X}$. Equation (4), in terms of the new variable $\boldsymbol{Y}$, becomes:

$$\mathrm{J} = \frac{1}{2}\boldsymbol{Y}^T B\boldsymbol{Y} + \frac{1}{2}[H(B\boldsymbol{Y}) - \boldsymbol{O}]^T R^{-1}[H(B\boldsymbol{Y}) - \boldsymbol{O}]. \quad (5)$$

Using the chain rule, the gradients of background and observation parts of the cost function (4) for **X** and cost function (5) for **Y** have the form:

$$\nabla_{\mathrm{X}} J = B^{-1}\boldsymbol{X} + H^T R^{-1}(H(\boldsymbol{X}) - \boldsymbol{O}). \quad (6)$$

$$\nabla_{\mathrm{Y}} J = B^{\mathrm{T}}\boldsymbol{Y} + B^{\mathrm{T}} H^T R^{-1}(H(B\boldsymbol{Y}) - \boldsymbol{O}) = B\nabla_{\mathrm{X}} J. \quad (7)$$

Equations (6) and (7) are simultaneously minimized by employing an iterative Conjugate Gradient process.

Start by assuming:

$$\boldsymbol{X}^0 = \boldsymbol{Y}^0 = 0. \quad (8)$$

Then iterate over n:

$$\nabla_X J^n = B^{-1}\boldsymbol{X}^{n-1} + H^T R^{-1}(H(\boldsymbol{X}^{n-1}) - O) = \boldsymbol{Y}^{n-1} + H^T R^{-1}(H(\boldsymbol{X}^{n-1}) - O), \quad (9)$$

$$\nabla_Y J^n = B\nabla_X J^n, \quad (10)$$

$$Dir \cdot \boldsymbol{X}^n = \nabla_Y J^n + \beta Dir \cdot \boldsymbol{X}^{n-1}, \quad (11)$$

$$Dir \cdot \boldsymbol{Y}^n = \nabla_X J^n + \beta Dir \cdot \boldsymbol{Y}^{n-1}, \quad (12)$$

$$\boldsymbol{X}^n = \boldsymbol{X}^{n-1} + \alpha Dir \cdot \boldsymbol{X}^n, \quad (13)$$

$$\boldsymbol{Y}^n = \boldsymbol{Y}^{n-1} + \alpha Dir \cdot \boldsymbol{Y}^n. \quad (14)$$

Dir is the direct transformation of the gradient of the original vector value. β is a norm of the gradient, and α is the final stepsize of iteration.

Until either the maximum number of iterations has been reached or the gradient is sufficiently minimized.

The final two equations are the actual value distribution from observational stepsize to model increment. X and Y are different representations of analysis increment, and the equation is directly related to the background error covariance. The whole-model grid and variable dimension information are saved on the background error covariance. From a single stepsize, the background error covariance distributes the increment to the whole three-dimensional grid and control variable (Hu et al. 2016).

Proper specification of the model background error is one of the most critical parts of the data assimilation technique, on which the performance of data assimilation depends highly. The National

Meteorological Center (NMC; Parrish and Derber 1992) method. The National Meteorological Center's spectral statistical-interpolation analysis system is used to obtain the background error covariance B, which extracts the systematic error by making the difference between the two model background forecasts started from the different initial times. In this study, the background error was calculated through the difference between the 24-hour forecast and the 12-hour forecast. The forecast differences were collected at 12-hour intervals to calculate the background error covariance statistics for May and June 2016 for the KORUS-AQ period and each season for January, April, July, and October 2017 for the All-Season case. The basic schematic of the NMC method has been described in Figure 2-4.

In this study, for sensitivity testing purposes, the EnKF data assimilation method and Meteorological Four-Dimensional Data Assimilation (FDDA) for WRF have been adapted.

EnKF method is one of the most generally adapted data assimilation methods, both for research and operational purposes. The basic concept of EnKF is generating probability distribution by the ensemble. EnKF simulates enough ensembles to generate the systematic bias from perturbed variants. The ensembles continuously simulate the three-dimensional data during the pre-designated periods, so unlike the 3DVAR, EnKF can generate the background error covariance more directly from their ensemble. Thanks to the ensembles, the EnKF method can generate the flow-dependent background error covariance. Each of the errors is representing a specific time, theoretically, EnKF can represent the temporal variation rather than the 3DVAR method.

EnKF is based on the Kalman filter algorithm, which is divided into the prediction step and observation update step. In the prediction step, evolve means the state is defined as below.

$$x^f = Mx^a. \tag{15}$$

x^f is the priori value, M is the forecast model, and x^a is the posterior value. From analysis, the background is generated from the forecast model. The relationship between the uncertainty of prior and posterior is defined as below.

$$P^f = MP^aM^T + Q. \tag{16}$$

P^f and P^a are uncertainty of prior and posterior values, and Q is the covariance of model error.

Next is the observation update step, and the posterior value is designated as below.

$$x^a = x^f + K\left(y - Hx^f\right). \tag{17}$$

H is the observation operator, and y is the observation value. K is a Kalman gain, which is defining how much the increment will be added, which is described as below.

$$K = P^f H^T \left(HP^f H^T + R\right)^{-1}. \quad (18)$$

R is observation error covariance. The updated uncertainty is computed as below.

$$P^a = (I - KH)P^f. \quad (19)$$

The ensemble approach includes ensemble is extending the assumption of Kalman filter, which was the model and observation operator is linear, to adapt the non-linear problem. An extended Kalman filter (Grewal and Andrews 2008) was developed to get around the problem, but it still needs an adjoint model to propagate the covariance matrix. Due to the size of the matrix, an alternate approach has been suggested that using the sample of evolved stats, which is ensemble.

The second is the FDDA method installed in the WRF model. FDDA is the nudging method to alternate the rough meteorological initializing. There are several options of FDDA, but in this study, analysis nudging has been adapted.

Analysis nudging is using the synoptic meteorological input to change the boundary condition and interpolates the meteorological variables grid by grid. Analysis nudging is related to the bias of grid analysis by input and model value. The forecasted equation for variable α is described as below.

$$\frac{\partial p^* \alpha}{\partial t} = F(\alpha, x, t) + G_\alpha \cdot W_\alpha \cdot \epsilon_\alpha(x) \cdot p^*(\hat{\alpha}_0 - \alpha). \quad (20)$$

F is the basic forcing term, and the other terms are nudging terms for p* and α. In the WRF model, p* is μ, which is pressure. F physical forcing term such as stream function or Coriolis force. Gα is nudging inverse time scale which defines the size of a term from the model process. Wα is vertical weight. εα is the quality control variable range between 0 and 1. ^α0 is the observational estimation from the model grid. The analysis grid nudging consists of 3D-analysis nudging and 2D-analysis nudging below the model PBL height (Stauffer et al. 1991).

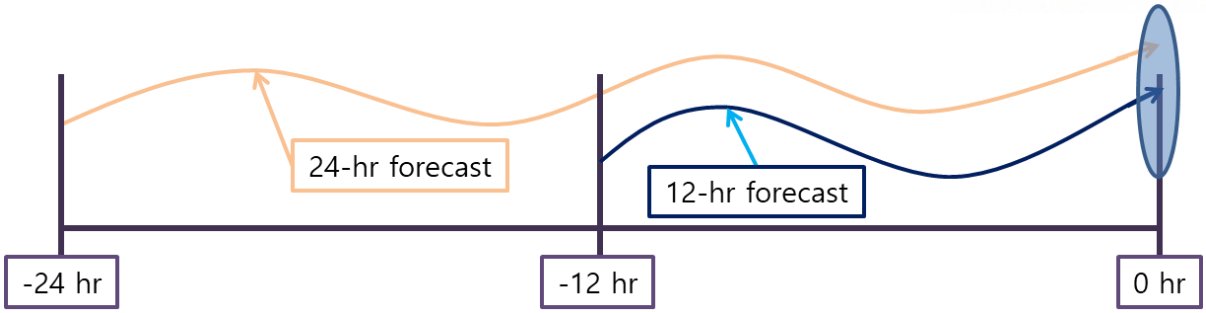


Figure 2-4. The concept schematic of the NMC method. The image specifies single background systematic error from 24-12 hour forecast biases from the different initial times.

2.3. Data

Both satellite AODs from MODIS and GOCI and the surface PM observations from the China Monitoring Network (CMN) and the National Ambient Monitoring Information System (NAMIS) in Korea were integrated for the input to the 3D-VAR data assimilation. Detailed information for these input data is provided in Table 1. It also shows the observation errors specified for the MODIS AOD as a function of the AOD value (τ), which follows Levy et al. (2013) for land (ε_L) and Remer et al. (2005) for the ocean (ε_O), respectively. The errors for GOCI follow the error equations specified by Choi et al. (2018).

Ground PM observations are available from 1,514 stations in China and 361 stations in South Korea, which are indicated in Figure 2-1. The PM observation error is represented with the measurement error and representative error. Measurement error is defined according to Schwartz et al. (2012), and the representative error is by Elbern et al. (2007) and Pagowski et al. (2010).

In addition, the quality of the observed data was controlled before data assimilation. For satellite AODs, only the data with a confidence flag 3, the best quality, were used. Data thinning was also applied where the high-resolution satellite data were coarsely sampled according to the model resolution (Liu and Rabier 2002). One AOD observation was randomly sampled for every 27 km by 27 km grid for Domain 1 and every 9 km by 9 km grid for Domain 2. In the case of PM_{10} ($PM_{2.5}$), the observation values were discarded as unrealistic when they were higher than 300 (150) $\mu g\ m^{-3}$ or the innovation (H (x) – y) values are higher than 300 (150) $\mu g\ m^{-3}$.

For the KORUS-AQ AOD analysis data verification purpose, AERONET is also included. AERONET is a sun-photometer installment to observe the atmospheric aerosol with light radiation, and it can generate AOD and Angstrom exponents for various wavelengths. There are 21 AERONET stations located in China, South Korea, and Japan during the KORUS-AQ period.

This study also uses the AOD reanalysis at 550 nm by the Modern-Era Retrospective analysis for Research and Applications version 2 produced by the National Aeronautics and Space Administration Goddard Space Flight Center (MERRA2; Gelaro et al. 2017). It provides a global analysis of aerosol produced by the analysis splitting technique (Randles et al. 2017) as the data assimilation method that applied to the Goddard Earth Observing System (GEOS) atmospheric model with the GOCART aerosol scheme. AOD observations include several sources, including the MODIS reflectance as the vast majority of the assimilated data since 2002. The MERRA2 provides both AOD analysis and chemical species categorized as GOCART.

2.4. Experimental Design

2.4.1 General Experiment Design

For the both KORUS-AQ and All-Season case, Data assimilation is conducted in a 6-hour interval. Each 00 UTC the meteorological conditions have been initialized by the UM analysis data. Because the geostationary GOCI data in visible radiation bands are only available hourly from 00 to 07 UTC, GOCI has been assimilated only for 00 UTC and 06 UTC. For surface in-situ data, 3DVAR data assimilation was performed at 00, 06, 12, 18 UTC for the aerosol fields to provide the initial conditions for the next 6 hours. The time window of GOCI has been set as ±3 hours to reduce the vacancy of observation data. This helps increase the number of satellite observation samples often masked by clouds. On the other hand, in-situ data only uses the exact period of data assimilation. The analysis data at 00 UTC every day is set as an initial condition for the WRF-Chem forecast for 24 hours. All the cycle-starting time is 00 UTC. For comparison, the NoDA run, a free WRF-Chem forecast run without starting from the analysis field obtained from the aerosol data assimilation but by restarting from the end of the previous 24-day run for chemistry and aerosol fields. In NoDA, the meteorological conditions are updated at 00 UTC every day as in the DA runs. Therefore, the difference between NoDA and DA is the impact of aerosol data assimilation. For both NoDA and DA, the chemistry fields were restarted from the previous forecasts. The schematic of data assimilation and forecast has been described in Figure 2-5.

Additionally, for the All-Season case, MODIS satellite AOD has been included for the data assimilation. The sun-synchronous MODIS data are available two times a day each at 03-04 and 17-18 UTC from Aqua, and 02-03 and 13-14 UTC from Terra, respectively. Due to the computational costs, only January, April, July, and October 2017 were selected for testing seasonal changes in the forecast performance, and for each month, the 24-h forecasts were conducted at 00, 06, 12, 18 UTC every day.

2.4.2 Sensitivity Test

For the comparison of the impact of the data assimilation variants, a sensitivity test has been conducted. Three major variants have been chosen, which are the data assimilation method, types of input observation, and fixing the meteorological condition. For the data assimilation method, the EnKF method has been selected. The EnKF test has been done for April 2017 to compare with the All-Season case. The basic data assimilation simulation process is the same as described in 2.4.1. 30 ensembles are

assigned and for perturbation purposes, emission inventory and meteorological schematics are perturbed. Meteorological schematic perturbation has been done for microphysics and planetary boundary layer schemes. Perturbation for each ensemble has been described in Table 2-1. For the sensitivity test, the primary EnKF test has been named EnKF_M, and the second test has been named EnKF_N. EnKF_N simulation is different from the EnKF_M, only difference does not initialize meteorological state at 00UTC.

For the type of input observation, GOCI and surface in-situ station data have been selected. The basic concept of this study is using multiple observations for one data assimilation, the sensitivity test contains single-observation assimilation categorized as satellite assimilation and in-situ assimilation. The experiment has been conducted for the KORUS-AQ period. Two additional experiments have been done to compare with the control KORUS-AQ experiment. DA1 is defined as satellite-only simulation, which only uses GOCI satellite AOD data. DA2 is defined as in-situ-only simulation, which only assimilates surface PM station data. These results are compared by DA3, which is a control simulation that assimilated both observations.

For the meteorological condition, the FDDA method has been selected. The method is the most basic method of gridded FDDA, which nudges the meteorological variables with input three-dimensional meteorological data. For this case, UM-GDAPS data has been nudged to the simulation. The nudging has been done for April 2017 and compared by 1 month. Additional FDDA experiment is FDDA, and control simulation is NF.

For the additional analysis of forecast data, a simple bias correction method has been suggested. The WRF-Chem forecast model generally performs underestimation of chemical concentration as the forecast duration becomes increased. To support the analysis, the bias correction has been done as the equation below.

$$X' = \left(X_f - \overline{X_f}\right)\left(\frac{\sigma_f}{\sigma_o}\right) + \overline{X_O}. \quad (21)$$

The X' is the corrected model forecasts, X_f is the original model forecasts, $\overline{X_f}$ *and* $\overline{X_O}$ are the time averages of X_f and observation, σ_f and σ_O are the standard deviations of the forecasts and the observations, respectively. The bias correction equation adjusts the model forecast by observational variations.

For the more sophisticated forecast skill comparison, the skill score has been adapted. A skill score is a quantified measure of the accuracy of prediction to an observation. There are several skill score calculation methods are existing, in this study, Heidke Skill Score (HSS; Heidke 1926). HSS is

calculated from the contingency table, the score is the number correct, or proportion correct. The contingency table for the WRF-Chem forecast is followed by the National Institute of Meteorology (NIER) in Korea, which divides the air quality into 4 categories. Each category are clean, normal, polluted, and extremely polluted, and the case for PM_{10} ($PM_{2.5}$) concentration, the ranges are 0-30 (0-15), 31-80 (16-35), 81-150 (36-75), and 151- (76-) ug m^{-3}. The HSS is computed for all cases and polluted cases, and its calculation is done by the equation below.

$$HSS = \frac{\sum_{i=1}^{I} p(y_i,o_i) - \sum_{i=1}^{I} p(y_i)p(o_i)}{1 - \sum_{i=1}^{I} p(y_i)p(o_i)}. \quad (22)$$

which i indicates the index of the category, I the total number of forecast categories, $p(y_i, o_i)$ the joint probability distribution of forecasts and observations, $p(y_i)$ and $p(o_i)$ the marginal distributions of the forecasts and the observations, respectively. $\sum_{i=1}^{I} p(y_i, o_i)$ represents the forecast proportion correct (i.e., the forecast accuracy of the actual forecasts), and $\sum_{i=1}^{I} p(y_i)p(o_i)$ proportion correct for the random reference forecast. Therefore, HSS indicates how much the forecast has been improved from the random reference forecast. HSS values range from negative to 1. The value of 1 means the perfect forecast, 0 indicates that the forecast has the same performance as the random forecast, and the negative values suggest even worse performance than the random forecast. When HSS is higher than 0, it indicates the forecast is practically useful.

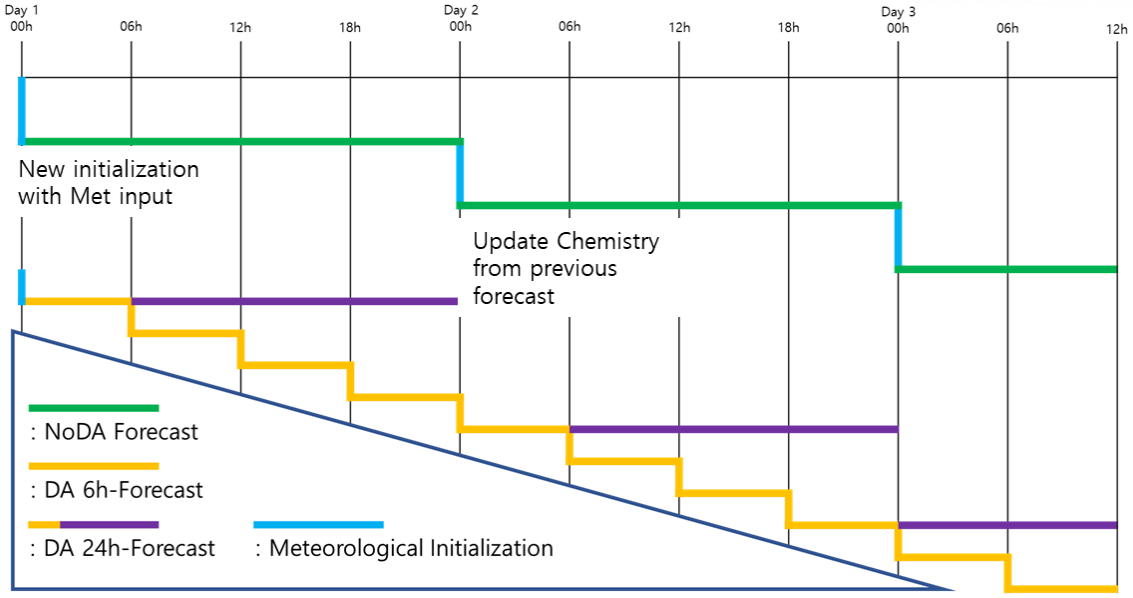


Figure 2-5. The schematic of how NoDA and DA/Forecast run will be done for the general simulation adapted to the both KORUS-AQ and All-Season case.

Table 2-1. Meteorological model perturbation information for each ensemble.

ENS	Microphysics	PBL
1-3		YSU
4-6		MYJ
7-9	Lin	Shin-Hong
10-12		MYNN2
13-15		UW
16-18		YSU
19-21		MYJ
22-24	WSM6	Shin-Hong
25-27		MYNN2
28-30		UW

3. Chapter 3

Results of Data Assimilation and Forecast

3.1. Verification of Data Assimilation Quality

3.1.1 Case Verification for KORUS-AQ

Figure 3-1 shows the mean NAMIS observation timely varying graph for the KORUS-AQ period. During the simulating period, there were three individual aerosol cases have been selected. Case A is the Asian Dust case (May 4 to 7), case B is the Local Pollution case (May 17 to 22), and case C is the Long-Range Transportation case (May 25 to 31). Case B and C are selected by Peterson et al. (2019) and Choi et al. (2021).

Asian Dust event in Case A is clearly shown in the observed PM_{10} concentration averaged over South Korea (Fig. 3-1a). The day of May 7 was officially recorded as the Asian Dust day. It was a clear day with no rainfall in South Korea with the daily maximum temperature ranging from 19 – 23 °C, characterized as typical springtime weather. The highest concentration is at 00 UTC on 7 May with a peak concentration of 200 μg m-3. There is also a build-up period of PM_{10} concentration during 11 – 23 May due to the accumulation of local air pollutants under the stagnant weather until washing out by rain on 24 May. There is another time of relatively high PM_{10} concentration during 25 May – 1 June. The ground PM_{10} concentration maintains at a high level above 80 μg m-3 on average, and this is attributed not to the Asian Dust event but the long-range transport from China.

The observed time series of $PM_{2.5}$ (Fig. 3-1b) show a similar temporal variation as in the PM_{10} time series, but also with interesting differences between PM_{10} and $PM_{2.5}$. Given that the PM_{10} concentration includes $PM_{2.5}$, the surface $PM_{2.5}$ contributes relatively more to PM_{10} in the long-range transport regime during 25 May – 1 June. It suggests the natural dust, mostly sand, consists of large particles, whereas the anthropogenic emission contributes more to the $PM_{2.5}$ component.

The observed AOD variation by AERONET stations (Fig. 3-1c) shows a somewhat different time variation to PM_{10} or $PM_{2.5}$. It does not show the highest values during the Asian dust event during 5 – 9 May. This implies that both variables do not always go along with each other. It suggests a weak coupling between the column-representing GOCI AOD and the surface-representing PM observations. Instead, the time series shows sporadic increasing events in the later period of KORUS-AQ, for example, around 25 May – 1 June and 7 – 11 June. The large temporal variation seems partly related to the statistical noise in the average with small data samples in the AERONET sites, reflecting high heterogeneity in AOD value in time and space.

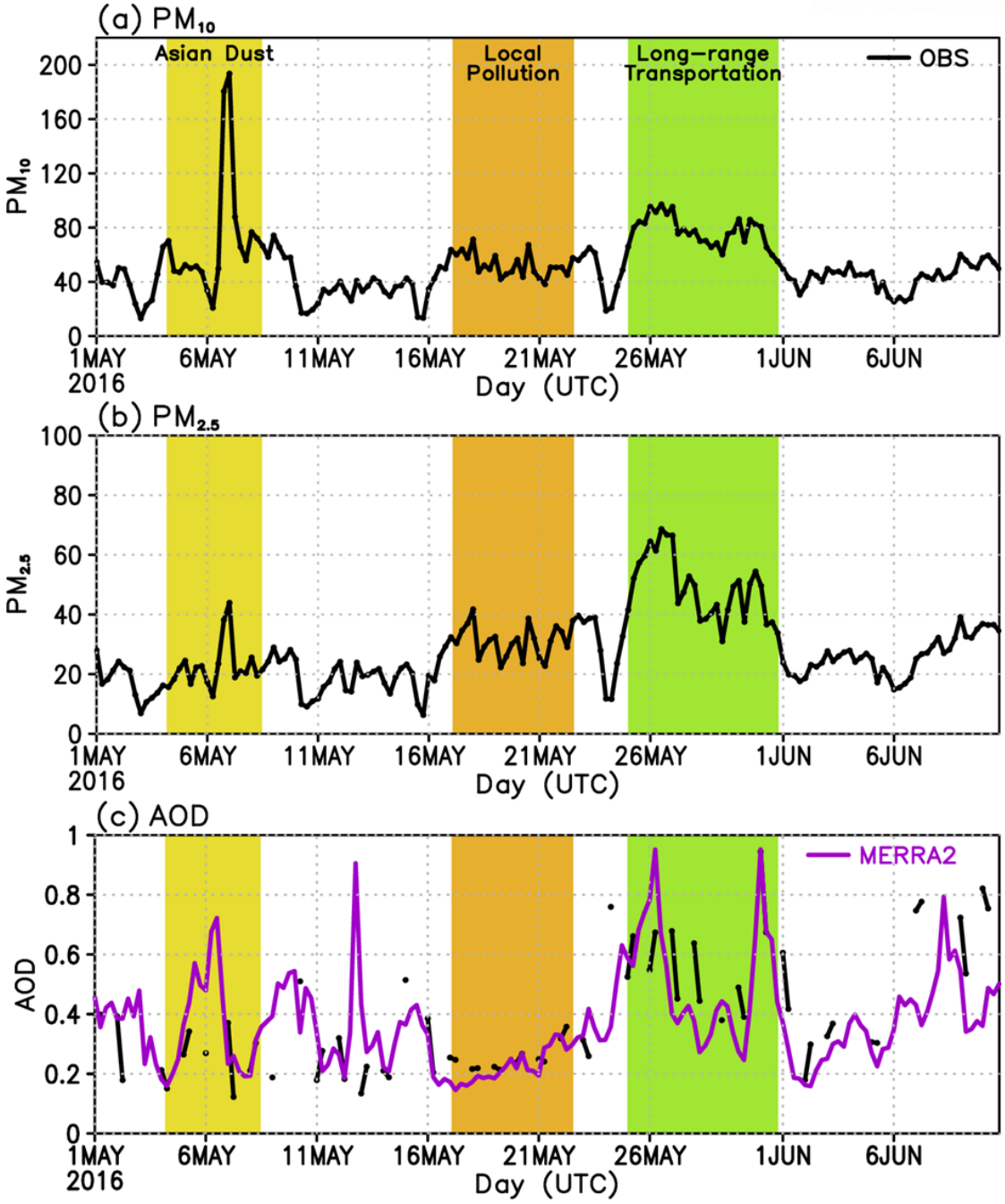


Figure 3-1. The timescale graph of observation for the KORUS-AQ period over the Korean peninsula. PM_{10} and $PM_{2.5}$ are collected from NAMIS, and AOD is collected from AERONET. For the AOD, MERRA2 reanalysis data has been added as reference data.

In the experiments, NoDA shows an underestimation of AOD as in the case of PM_{10} or $PM_{2.5}$, although it follows the observation time series with much smoothed temporal variation. The NoDA time variation is overall weaker than the observed. The model shows a limitation in detecting high-frequency variability and the high concentration episodes when no observational data is provided, presumably due to the model deficiencies in emission and transport processes. The simulations of DA are more consistent and show a better consistency with the AERONET and In-situ PM observations.

DA shows the best fit to the observed time series for the entire period of KORUS-AQ. Especially in the Asian dust period on 7 May, local pollution period of 18 May – 22 May, and the long-range transportation period of 25 May – 1 June, it reproduces observed PM_{10} concentrations properly. The results highlight that utilizing multiple observation sources, including satellite and ground observation data, can attain the best-fit analysis. DA experiment reproduced the observed $PM_{2.5}$ concentration remarkably well. This is again due to the direct use of surface $PM_{2.5}$ observation in the data assimilation process, which tends to correct the systematic underestimation bias in NoDA as in PM_{10}. For the AOD case, the DA result shows a relatively good relationship to the AERONET result. Although for the later stage, DA and AERONET observations do not fit well but compared to the MERRA2 reference, the overall temporal variance is fitted well. The control run for KORUS-AQ showed significantly reliable to the PM and AOD more than the background model.

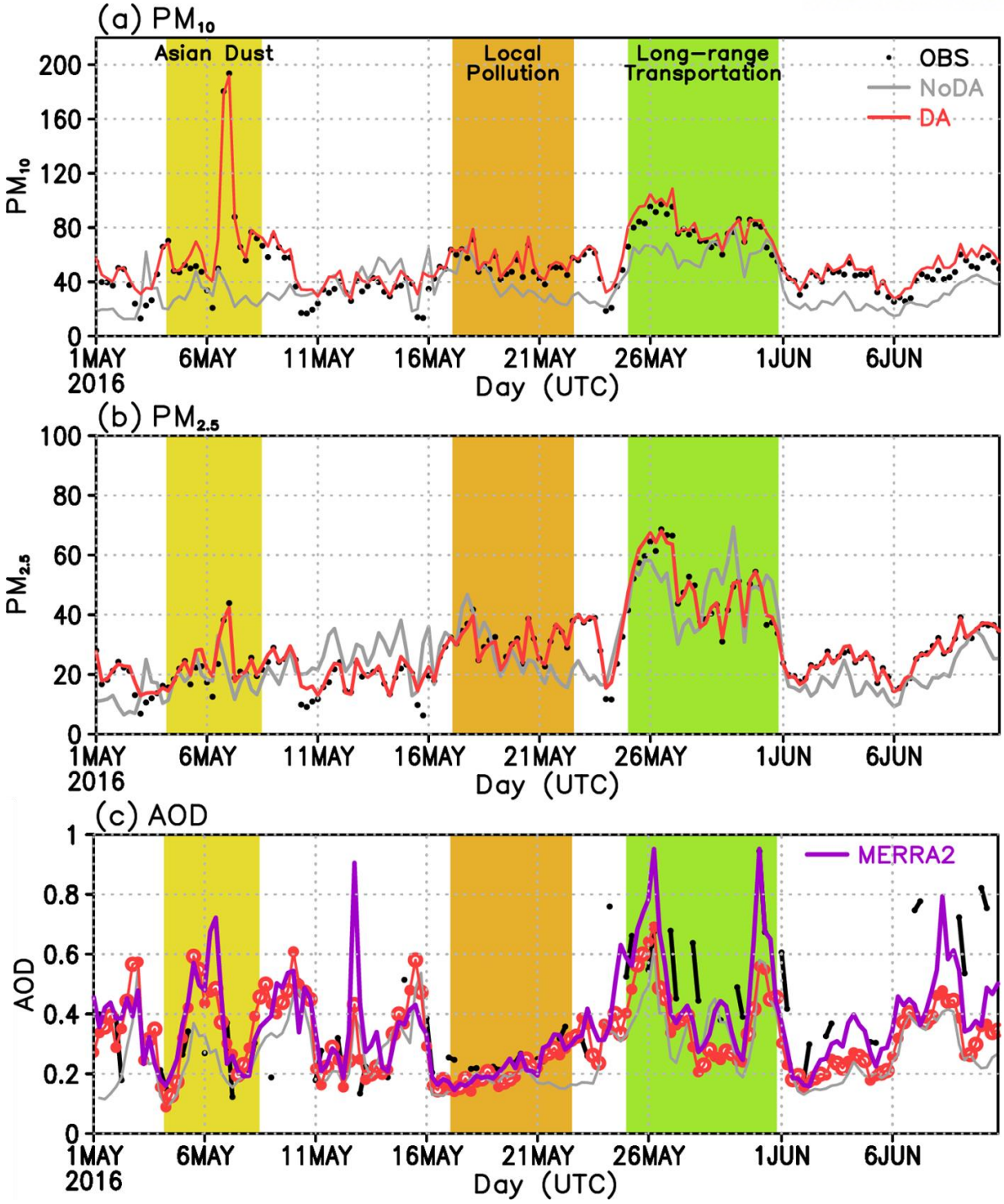


Figure 3-2. The timescale graph of observation and DA results for the KORUS-AQ period over the Korean peninsula. PM_{10} and $PM_{2.5}$ are collected from NAMIS, and AOD is collected from AERONET. For the AOD, MERRA2 reanalysis data has been added as reference data.

For the detailed analysis, three cases of KORUS-AQ have been analyzed independently to validate each data assimilation with the GOCI AOD and ground PM observations. First is the Asian dust event that occurred on 7 May 2016 (Fig. 3-3). Although the GOCI observations show many missing values over the domain due to the signal contamination by clouds, they give a hint of the AOD plume of high concentration levels across the southern part of the Korean Peninsula. It also shows the high AOD values in the East Sea. The model background (NoDA) simulates the dust plume that originated from China and was transported to South Korea and Japan (Fig. 3-3c). However, the simulated AOD values are much weaker than the GOCI AOD, even though the WRF-Chem model driven by the synoptic-scale weather patterns can simulate the dust event qualitatively in Northeast Asia. In South Korea, NoDA tends to simulate relatively high AOD values in the south but not as high as in the GOCI observations. It also represents the high AOD values in the East Sea realistically (but with somewhat weaker magnitudes than GOCI), which is not featured at all in NoDA. An AOD increase but in weak magnitude is also found in South Korea. Compared to the MERRA2 result, MERRA2 shows even higher AOD than GOCI for some regions, but the aerosols front pattern over the Korean Peninsula is similar to the DA result. The assimilation results with the ground PM observations only show that the dust plume becomes much weaker over the oceans and in Japan, where no ground observations are available. This suggests a significant beneficial impact of using satellite AOD values in the aerosol data assimilation that fills the gap of ground observations. When both GOCI AOD and ground PM observations are assimilated altogether.

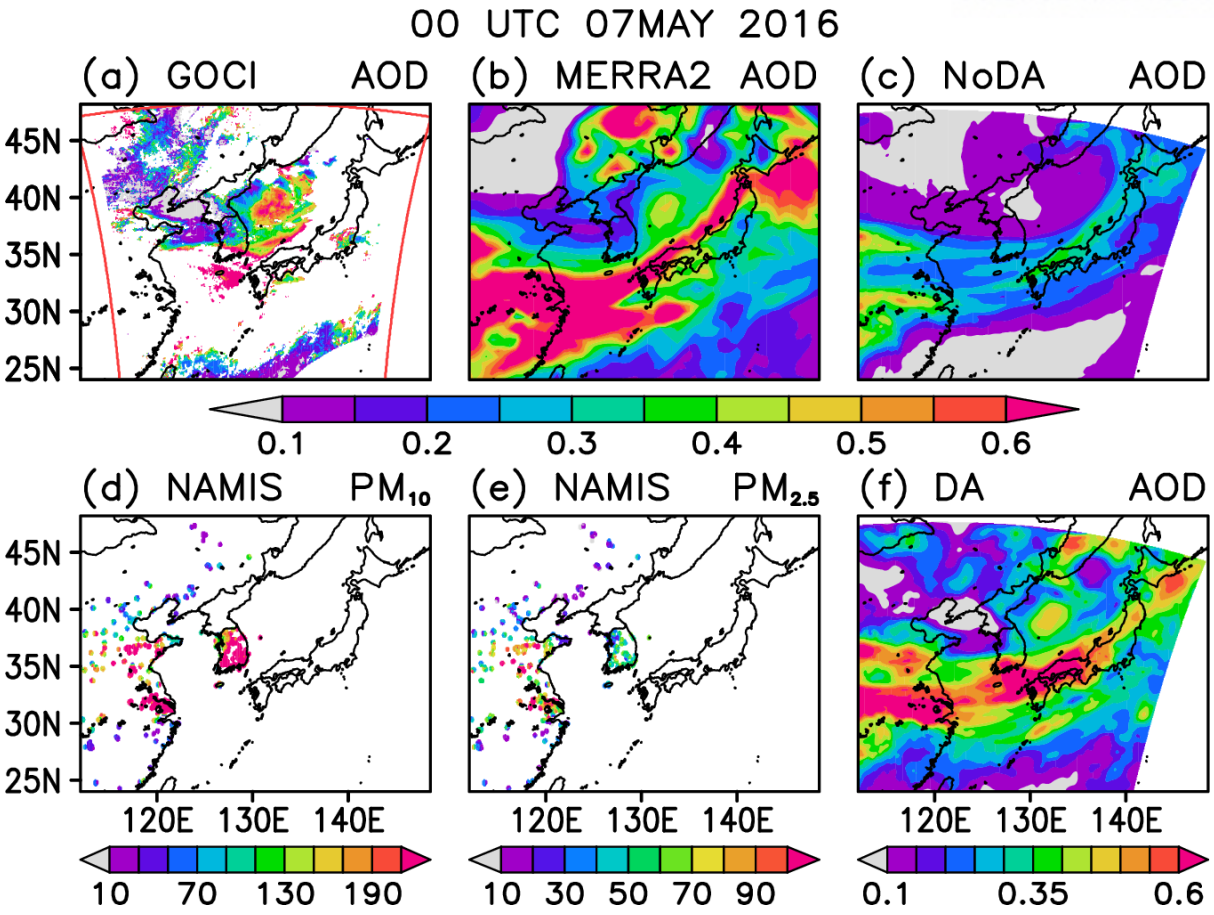


Figure 3-3. The AOD distribution at 00 UTC on 7 May 2016 over Northeast Asia from (a) the GOCI observations, (b) MERRA2, (c) NoDA (with no assimilated observations), (d) NAMIS-PM_{10}, (e) NAMIS-$PM_{2.5}$, and (f) DA (GOCI and ground PM). The area within the red frame in (a) indicates the spatial coverage by the GOCI instrument. The unit of AOD is dimensionless. The unit of PM is ug m^{-3}.

For case B, the local pollution pattern showed a different pattern compared to case A. In GOCI observation (Fig. 3-4a), the surroundings of the Korean Peninsula show higher contamination, and the high aerosol mass is located over Japan, trailed to Korea. From the referencing MERRA2 (Fig. 3-4b), the overall AOD pattern shows wider regions than GOCI, especially over the China and southern part of the domain. Both GOCI and MERRA2 show highly polluted areas over the northern part of Japan and the Manju region. Comparing the NoDA (Fig. 3-4c) and DA (Fig. 3-4f), the high aerosol pattern over Japan showed a clear difference which the original background model has not been simulated. DA result shows fill the high concentration region well for both Manju and Japan region. Compared to the MERRA2 and GOCI, the North Korea region shows opposite AOD, which MODIS satellite did not catch up properly. The DA result reflected the AOD over Korea Peninsula reliable to the GOCI AOD. The other different regions are the Shandong peninsula and the Manju region, the GOCI AOD showed relatively high AOD, but not for the MERRA2. The Shandong peninsula showed exclusively lower AOD than GOCI, and the Manju region seems that the high-AOD region is located more northward region than GOCI. For the Manju region, GOCI might be showing some noises from the cloud because of the excluded data for the center region and high AOD for surroundings, but it can be another discrepancy from two AOD observation installments. GOCI AOD has been directly adapted to the DA result, and the DA result is showing a similar pattern for GOCI rather than MERRA2 over the Manju region. Other than that, most of the AOD results are reliable to both of the references. Compared to the surface PM data (Fig. 3-4d and e), the surface concentration pattern of PMs over South Korea seems similar to the AOD. The PM concentration over China showed that the JJJ region showed the highest concentration and DA showed somewhat similarity, but for the Manju region, surface concentration seems much smaller than column concentration.

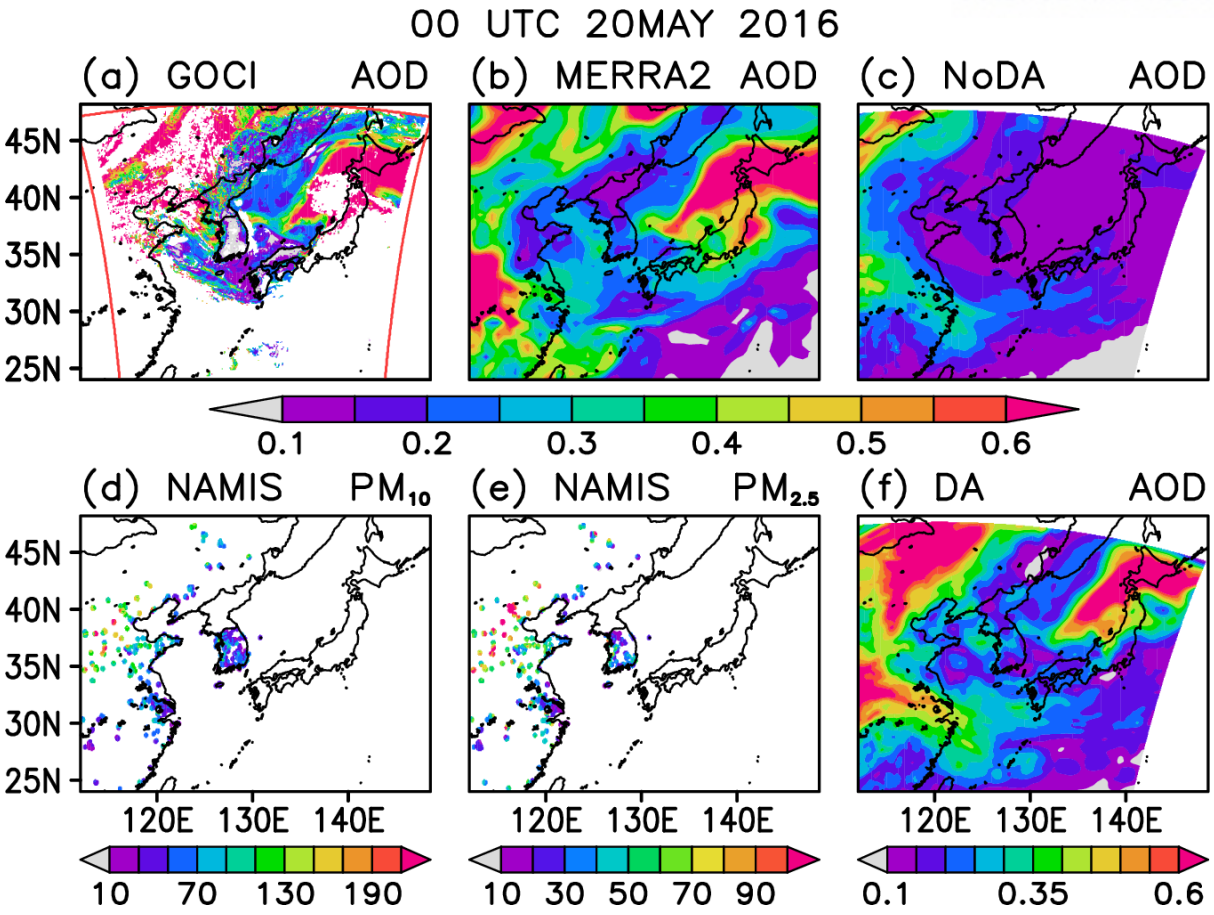


Figure 3-4. The AOD distribution at 00 UTC on 20 May 2016 over Northeast Asia from (a) the GOCI observations, (b) MERRA2, (c) NoDA (with no assimilated observations), (d) NAMIS-PM_{10}, (e) NAMIS-$PM_{2.5}$, and (f) DA (GOCI and ground PM). The area within the red frame in (a) indicates the spatial coverage by the GOCI instrument. The unit of AOD is dimensionless. The unit of PM is ug m^{-3}.

For case C, the long-range transport event in South Korea from China is shown in Fig 3-5. In the GOCI observations (Fig. 3-5a), AOD indicates high values from 35N of the west to 40N of the east, although most of the satellite data are contaminated by the noise. NoDA (Fig. 3-5c) simulates the high AOD plume stretching from the inner lands of China to the Hokkaido in Japan through the middle region of the Korean Peninsula. Although the model experiment with no data assimilation happens to reproduce the high AOD values, the transport path tends to shift to the north so that the AOD values in South Korea become weak. In DA (Fig. 3-5f), it improves the magnitude of AOD plumes and expands the plume by north and south, and covers most of South Korea with high AOD values. The satellite data assimilation helps improve the values in the ocean points and the downstream part, such as in the northern part of the main island of Japan. DA result picked up the GOCI signal very well because most of the GOCI signals are removed and linear transporting patterns are nearly invisible in the GOCI results. There two more different aerosol contaminated regions between MERRA2 and DA results, which are the region near the Slavyanka and the inland of China. The case of Slavyanka is that the GOCI and MERRA2 show opposite results, GOCI only shows the high-aerosol pattern more northward region. For China, almost all of the GOCI observations are missing. Analyzing the MERRA2 result, inland China shows another unique AOD pattern. This pattern is connected to the linear transporting pattern and seems like aerosol from China is transported eastward to Japan. This pattern of aerosol is weaker in DA results because of the weaker data assimilation effect over inland China. For surface PM, the highest PM concentration is located near the Shandong peninsula. AOD pattern is tilted to the southward direction and this high AOD seems not much related to the surface PM, most of the transporting aerosols are located at the upper altitude rather than the surface.

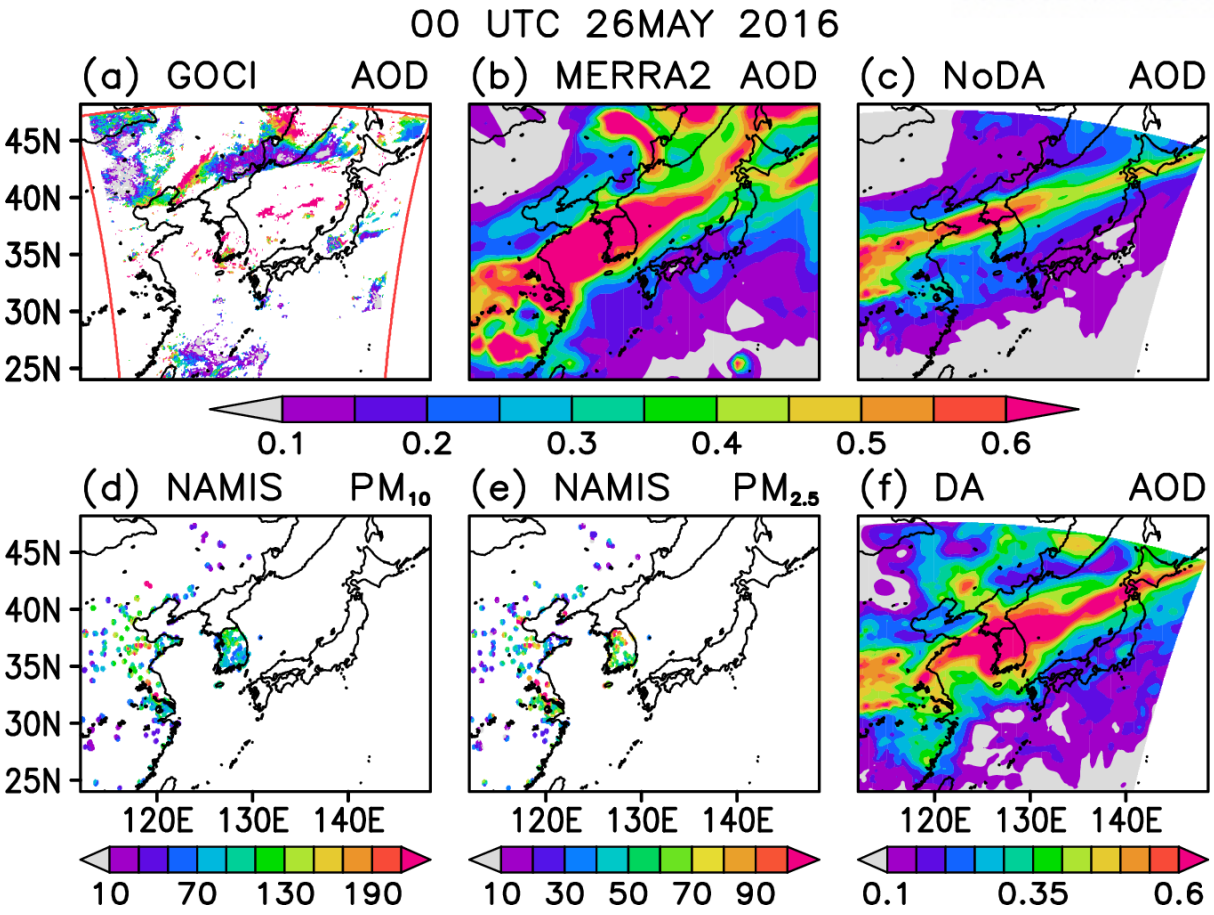


Figure 3-5. The AOD distribution at 00 UTC on 26 May 2016 over Northeast Asia from (a) the GOCI observations, (b) MERRA2, (c) NoDA (with no assimilated observations), (d) NAMIS-PM_{10}, (e) NAMIS-$PM_{2.5}$, and (f) DA (GOCI and ground PM). The area within the red frame in (a) indicates the spatial coverage by the GOCI instrument. The unit of AOD is dimensionless. The unit of PM is ug m^{-3}.

3.1.2 Case Verification for All-Season

Figure 3-6 shows the average ground PM_{10} concentration for Domain 1 and 2. The 6-hourly PM analysis fields are averaged for January, April, July, and October in 2017 to represent four seasons. The values over the desert area are masked out intentionally in the figure due to excessive dust concentration in the model but with no reliable observational data to verify. The spatial patterns of both domains are quite similar in NoDA (c.f. Fig. 3-6b and 3-6f). In domain 1, NoDA underestimates the concentration in Manchuria and the Korean peninsula and overestimates over the Sichuan basin and the inland areas of China. In domain 2, South Korea showed the total underestimation. DA shows much improvement when compared with NoDA. First, the overall overestimation bias in the middle and southern China and the underestimation bias in northern China and the Korean peninsula in the PM_{10} concentration is reduced through the data assimilation. In particular, the concentration has increased at 40° N latitude belts from BTH to Korea, Japan, and the Pacific Ocean.

Figure 3-7 compares the spatial distribution of the time-averaged $PM_{2.5}$ concentration for Domain 1 and 2. Most of the results are similar to Figure 3-6, but the actual increment pattern shows smaller and the locality pattern is more significant than the PM_{10} case.

Compared to Figures 3-6 and 3-7, the PM_{10} and $PM_{2.5}$ patterns are similar to each other. As in East Asia (Domain 1), NoDA significantly underestimates the PM_{10} and $PM_{2.5}$ concentrations over South Korea. The only exception is the Beijing-Tianjin-Hebei (BTH) area, known for the most polluted region in China, where NoDA underestimates PM_{10} but overestimates $PM_{2.5}$. For Domain 2, surface PM concentration shows quite inhomogeneous spatial patterns with the maxima in the northwest around the Seoul Metropolitan Area and its vicinity. As the surface observation sites are clustered mostly in the cities, the PM concentration also shows local maxima around those cities but with many gaps of data over an unobserved rural area.

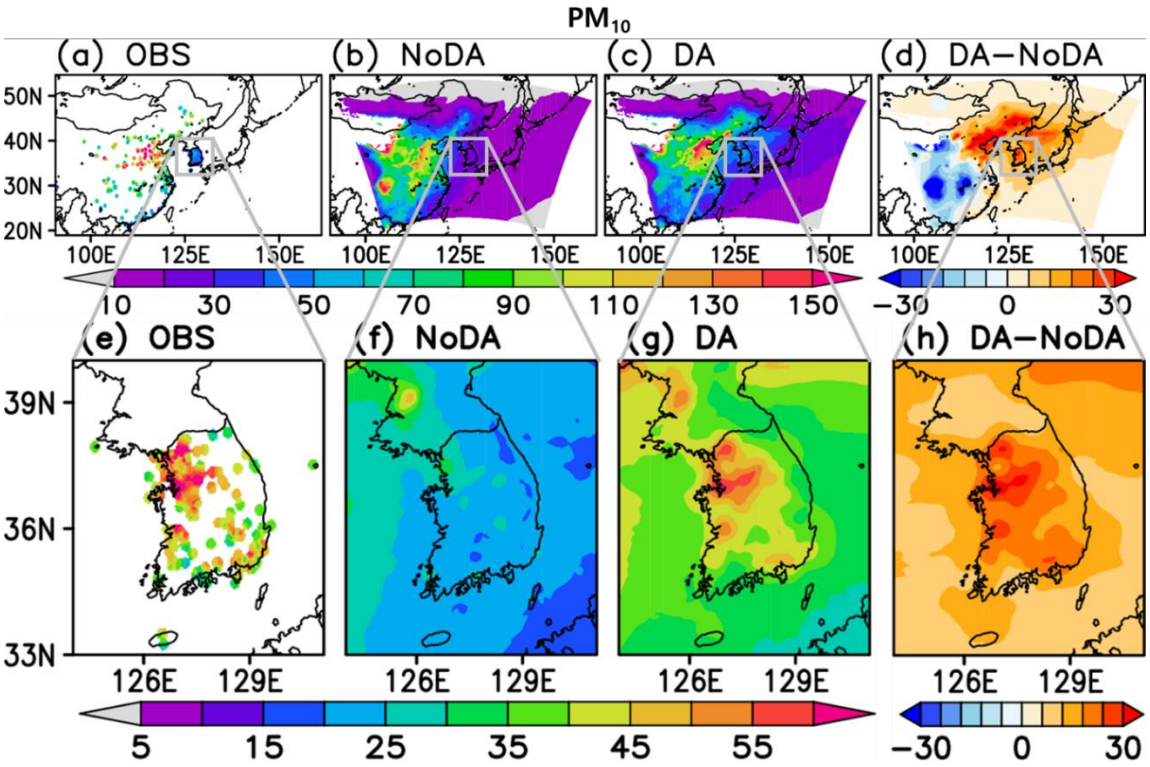


Figure 3-6. Seasonal average of surface PM_{10} concentration analysis at the lowest model level for the (a, e) Observation, (d, f) NoDA, and (c, g) DA experiments on Domain 1. (d, h) show the differences of DA and NoDA experiments. The top figures are showing domain 1 and the bottom figures are showing domain 2. The desert region is masked out.

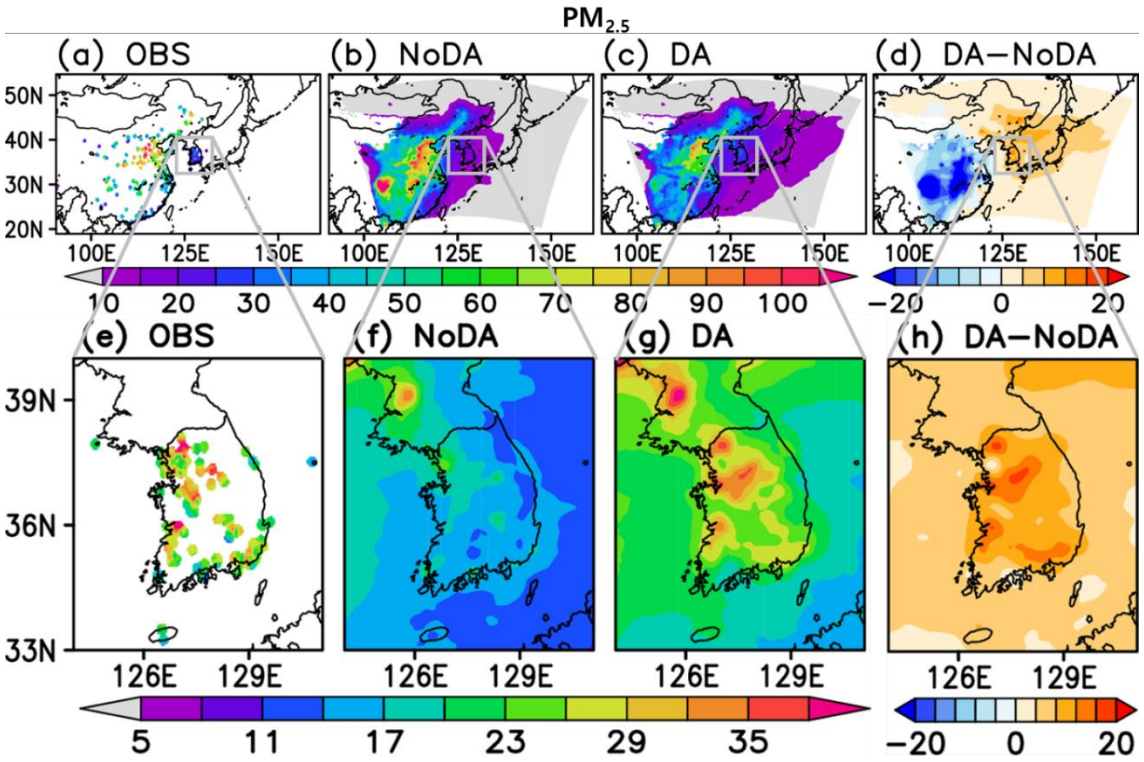


Figure 3-7. Seasonal average of surface $PM_{2.5}$ concentration analysis at the lowest model level for the (a, e) Observation, (d, f) NoDA, and (c, g) DA experiments on Domain 1. (d, h) show the differences of DA and NoDA experiments. The top figures are showing domain 1 and the bottom figures are showing domain 2. The desert region is masked out.

Figure 3-8 shows the scatter plots from surface PM concentrations to verify the simulation results at the observation sites. The data has been collected from 4 months of each season. When comparing the scatter plots, the scatter plots of the NoDA experiment are uneven in the overall shape of both PM_{10} and $PM_{2.5}$, and the models are generally underestimating the PM concentrations than observations. It shows low correlation and high RMSE. Especially, in the case of correlation, PM_{10} is 0.34 and $PM_{2.5}$ is 0.44, showing low consistency with the observed data. Compared to the irregular scatter plot of NoDA, the DA result was highly fitted and aligned on the diagonal line very well, and the correlation was very high as 0.86 for PM_{10} and 0.88 for $PM_{2.5}$, and it was verified that the data assimilation effect was high. In addition, it can be seen that RMSE is also significantly reduced. From these results, it can be seen that the PM concentration in the analysis site was effectively improved through the data assimilation system. The result contains 4-month analysis results. From the 4-month samples, the common data assimilation skill showed the assimilation skills of significantly reliable to the observation.

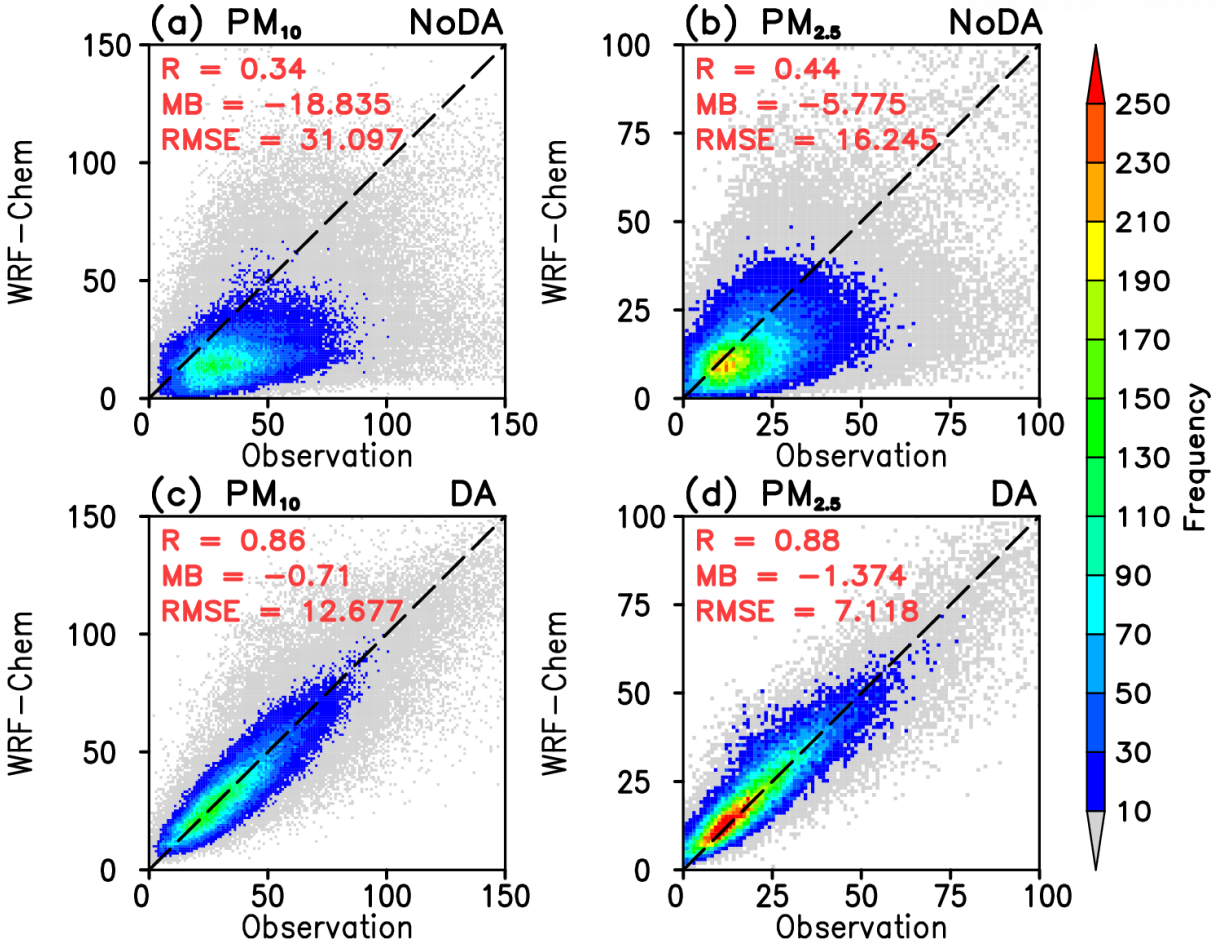


Figure 3-8. Scatter plots of PM_{10} and $PM_{2.5}$ between observations and the model analysis from NoDA and DA experiments over South Korea. The shaded values indicate the frequency. The correlation (R), mean bias (MB), and the root-mean-squared error (RMSE) values are also provided in each panel.

Figure 3-9 compares the vertical distributions of the standard deviation of the model background error averaged for one month and each species of GOCART. The errors are generally large in the lower atmosphere as most of the aerosols are originated from the surface. Large concentration change in the near-surface levels contributes to the model errors and forecast uncertainty. The background error statistics depend much on the season because the emission sources and the dominant atmospheric circulation patterns responsible for aerosol transport are quite distinctive across the season in East Asia. On average, the errors tend to increase in winter by organic carbon, $PM_{2.5}$, and black carbon, in descending order, which suggests increased fossil fuel combustion and biomass burning in this season. Stronger surface wind by enhanced East Asian jet stream also contributes to more active aerosol transport and large forecast uncertainty for these species. Hydrophilic aerosols (BC1 and BC2) exhibit more uniform errors within the atmospheric boundary layer up to 800 hPa, while the hydrophobic (BC1 and OC1) ones show a much rapid decrease from the surface. For those three species, the rest of the seasons exhibit comparable error magnitudes.

The natural dust error becomes maximum in spring due to more frequent Asian dust transport from the inland continental area. The mid-troposphere near the 600 hPa level shows the largest forecast error instead of the surface, implying that this species is being transferred by background winds rather than uploaded from the surface in Domain 2. When the prevailing wind in the low level changes to southerly or southeasterly from the ocean toward the continent in summer, the error for dust shows the largest values in the upper troposphere where the westerlies are prevailing. In contrast, the forecast error for natural sea salt does not show clear seasonal variation.

The background error for sulfate is the largest in summer and the smallest in winter. Sulfate aerosols are produced mostly by anthropogenic sources such as power plants and industrial processes. In our model configuration, WRF-Chem is forced by the standard emission inventory of EDGAR-HTAP, which provides the sulfur dioxide (SO_2) gas emission only, and there is no emission in sulfate aerosols. Therefore, all sulfate aerosols in the model forecast are formed secondarily. In our investigation, even though the SO_2 emission by EDGAR-HTAP peaks in winter, the simulated sulfate aerosol concentration peaks in summer. The transformation of SO_2 to the sulfate aerosols is affected by many factors including temperature, photolysis, and other meteorological conditions. The summer maximum in sulfate concentration in the model is also supported by the observational data in Seoul.

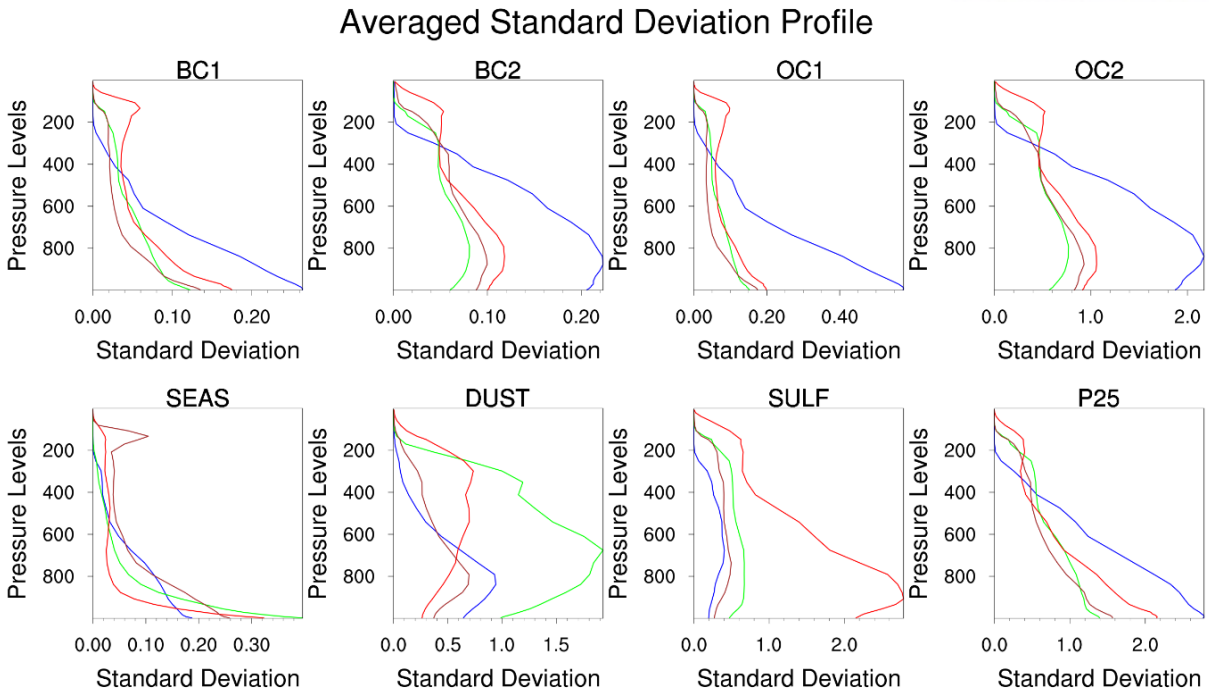


Figure 3-9. Vertical distributions of the standard deviation of the model background error for 8 GOCART aerosol species averaged over Domain 2 in winter(blue), spring (green), summer (red), and autumn (brown). Values in four different size bins are averaged for dust and sea salt.

Figure 3-10 shows the average concentration and standard deviation of PM by season in Korea. In Korea, January is winter, April is spring, July is summer, and October is autumn. Looking at the observed value of $PM_{2.5}$, it has a seasonal pattern with the highest concentration in January and then low concentration values in April, July, and October. This is consistent with previous studies. It has a high PM concentration due to dry and stagnant weather conditions in winter and spring and the lowest concentration in summer due to the washout effect of East Asian monsoon precipitation (Lee et al. 1999). In the case of PM_{10}, it is similar to $PM_{2.5}$ but has the highest concentration in April, which is due to the Asian Dust flying from the continent in spring (Jung et al. 2019; Kim et al. 2016). The NoDA experiment tended to underestimate the seasonal mean concentration of PM_{10} and $PM_{2.5}$ and have a smaller standard deviation than the observation. The DA experiment showed that the average concentration for each season was very similar to the observation, and the overall standard deviation was well matched. When the NoDA and DA experiments were compared, the difference in concentration and standard deviation were visible, and through data assimilation, the quality of the analysis site was improved, and the seasonality of Korea was well simulated. In addition, the statistics for each season has been shown in Fig 3-10c, d, e, and f. Correlation shows before and after the data assimilation, and the result shows that the April case consists of the most improvement by data assimilation, because of the bias between NoDA and DA. Compare to the other seasons, the correlation change of other seasons are similar to each other, but not for April. It showed that the April case consists of somewhat significantly different events or aerosol concentrations than other cases, and that case is improved properly after the DA. For RMSE, on the other hand, the bias between NoDA and DA seems different pattern than correlation. The January case shows the highest improvement based on RMSE, and the October case is the least. RMSE focuses more on the actual difference of the concentrations and January is the best-simulated case even the ambient concentration is relatively high which the NoDA is high. Generally, correlation and RMSE have a clear negative relationship, but in this study, the relationship is not much confident for the All-Season case.

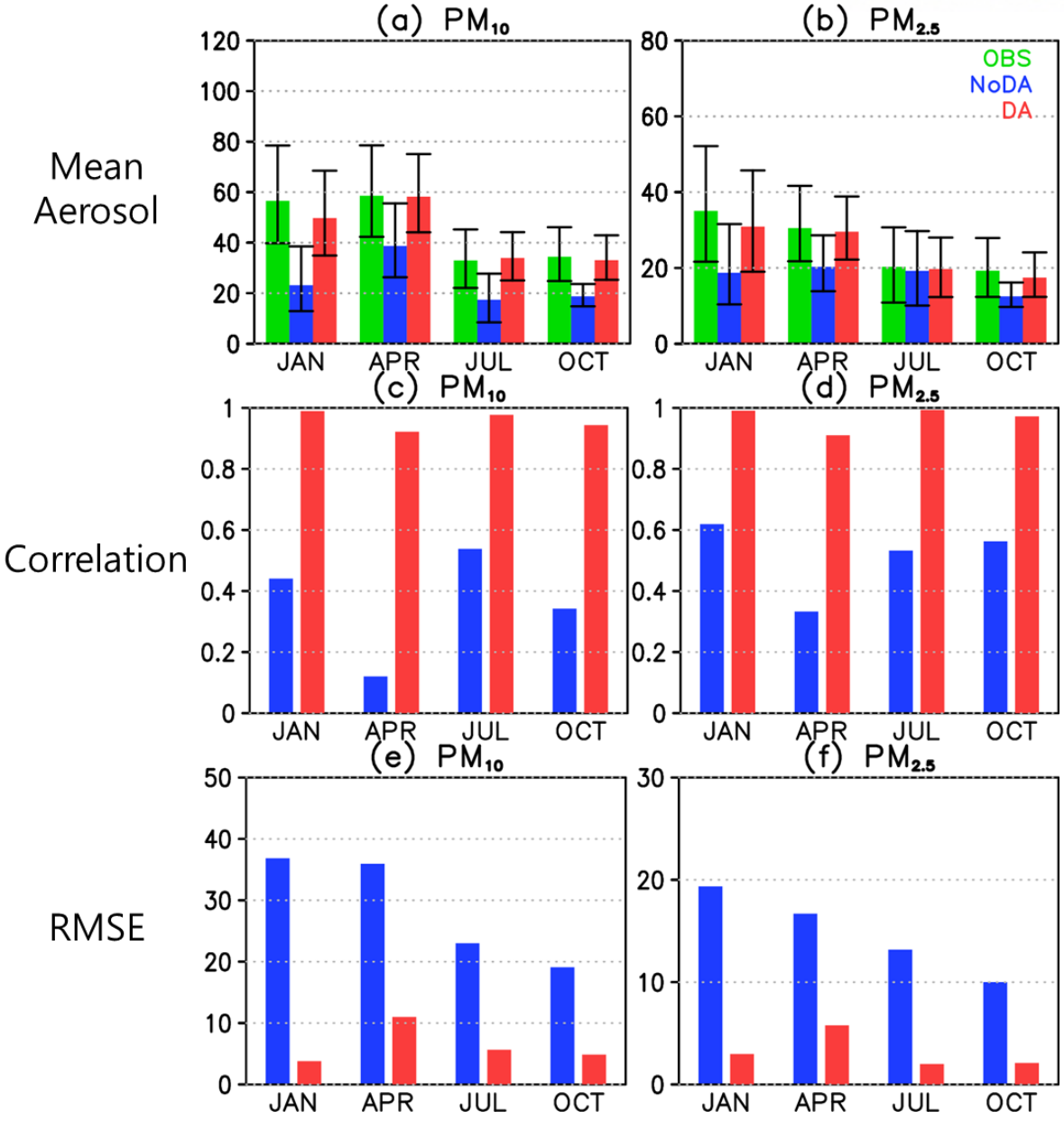


Figure 3-10. The analysis-averaged (a, c, e) PM_{10} and (b, d, f) $PM_{2.5}$ concentration and statistics in South Korea each season from the observations (green), NoDA (blue), and DA (red). The one-standard deviation of the daily values is indicated in each bar for (a) and (b). (a) and (b) shows mean aerosol, (c) and (d) shows the correlation, (e) and (f) shows the RMSE.

Finally, to confirm the effect of data assimilation for each season and to analyze where the average rate of change of concentration is large, the increment for each season is shown in Fig 3-11. As a result of seasonal PM increment comparison in Korea, the concentration was increased in winter and summer, and the change in concentration was relatively not significant in autumn. In the case of spring, it was positive in some regions, but negative increment was dominant from the only one among the four seasons. When comparing the increments for each aerosol type, the negative increment of dust was the largest (not shown), suggesting that sand dust such as Asian dust in the spring was overestimated and improved toward reducing after data assimilation. The increment pattern of the sum of all dust patterns is significantly similar to the PM_{10} increment of the spring (Fig. 3-11b), which means that the PM_{10} for spring is dominated by the dust transportation from the west. The $PM_{2.5}$ pattern is not much related to the dust increment, which the spring season is mostly affected by the Asian dust phenomenon. The period when the increment in Korea changes the most is in winter, where the increment in local concentration is dominant, and there are quite steep concentration changes between adjacent regions. This suggests that data assimilation, not a model, is essential for the appropriate adjustment of the analysis. Ultimately, for all seasons, spring is affected most by atmospheric transportations, winter is dominated by the local emission, summer shows the underestimation of aerosol and autumn shows the least change between background and analysis.

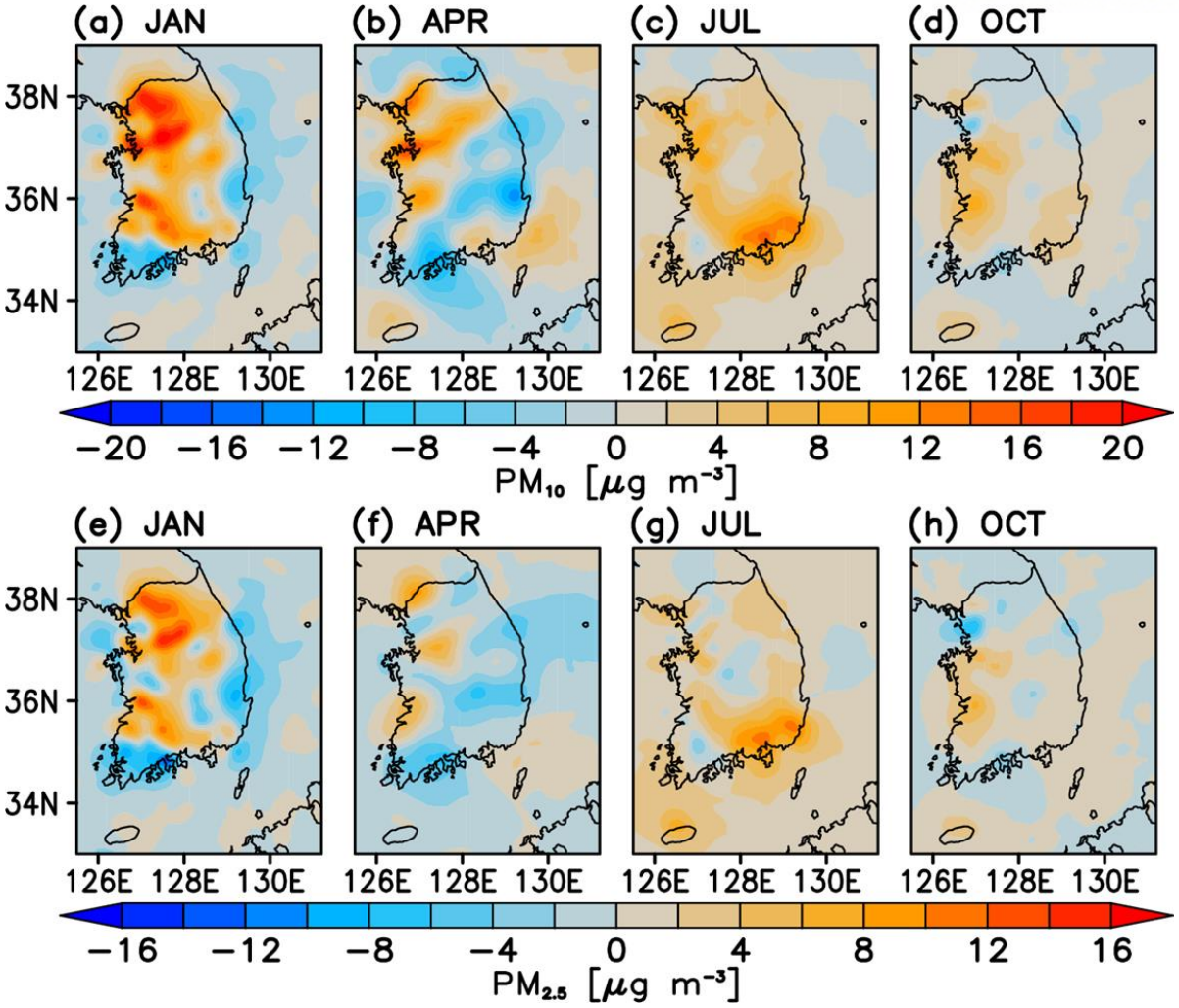


Figure 3-11. Monthly mean increments of data assimilation for each season over South Korea. (a, b, c, d) PM_{10} and (e, f, g, h) $PM_{2.5}$.

3.2. Impacts on Air Quality Forecasts

3.2.1 Forecast Analysis for KORUS-AQ

For the Asian dust case, comparing Figure 3-12c and 3-12d, 24-hour simulation maintains the high-AOD fronts over the southern region of the peninsula. Compared to Figure 3-12e, actual AOD has been decreased, and compared to the GOCI, the high AOD over the East Sea is not visible anymore like NoDA, the data assimilation effect has remained even after 24 hours of the previous assimilation performance. Asian dust front has remained and the highest aerosol-contaminated regions are also visible compared to MERRA2. The Asian dust case is one of the unique cases that the pure background model that has been properly provoked the dust from the Gobi desert and transported towards Korea and Japan. The 24-hour forecast only shows a synoptic frontal pattern which is shown similar to the NoDA simulation, and local highs over the East Sea and Hokkaido are not properly resolved. The results showed that the 24-hour forecast can maintain the data assimilation effect only for the synoptic phenomena and patterns which is continued from the last simulation, and it is hard to maintain the local patterns and temporal phenomena only can be observed in the observation.

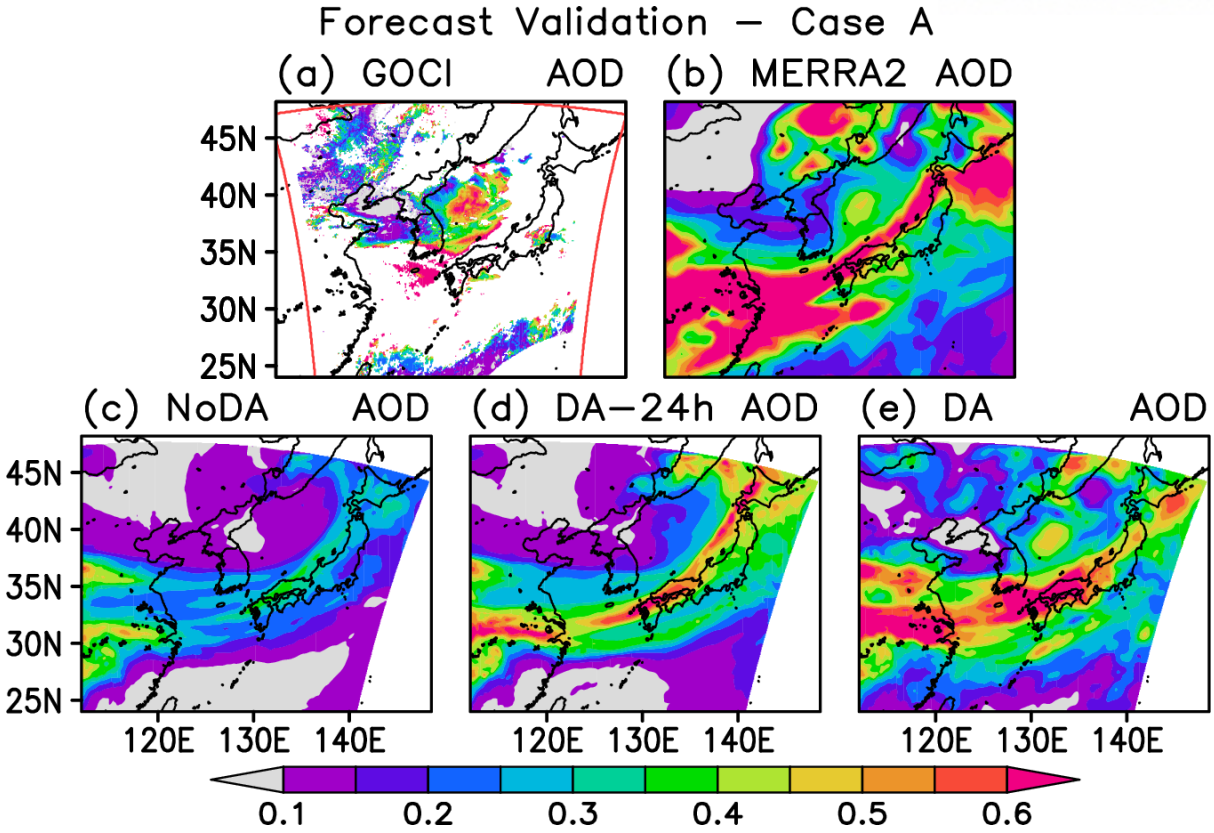


Figure 3-12. The AOD distribution at 00 UTC on 7 May 2016 over Northeast Asia from (a) the GOCI observations, (b) NoDA (with no assimilated observations), (c) DA1 (the assimilation of GOCI only), (d) DA2 (ground PM only), and (e) DA3 (GOCI and ground PM). The area within the red frame in (a) indicates the spatial coverage by the GOCI instrument. The unit of AOD is dimensionless.

For the local pollution case, comparing Figure 3-13c and 3-13d, the significant difference is the remaining high AOD over Hokkaido, Japan in 24-h forecast. NoDA cannot resolve the AOD over Japan, but a 24-hour forecast from previous data assimilation resolves the surrounding patterns. China region and southern sea showed not much difference between NoDA and the forecast, the data assimilation can fill the void section of the background model, which is not resolved properly by NoDA. The unique characteristics of the local pollution case are that most of the highly polluted regions have remained after the 24-hour simulation. This means that the meteorological conditions of the current domain for this case are quite stable and the atmospheric transportation is stagnated, which the data assimilation effect can be maximized. Especially, the high AOD patterns over Hokkaido and Japan regions are not present in the NoDA result, the data assimilation procedure is crucial for this period to maintain the synoptic aerosol patterns. The 24-hour simulation patterns over the Manju regions are much similar to the MERRA2 than GOCI, it seems that the MODIS observation is not able the observe the proper AOD pattern over the Manju region at the proper time. This might show the limitations of the polar-orbit satellite, which is hard to observe at a certain point at the demanded time.

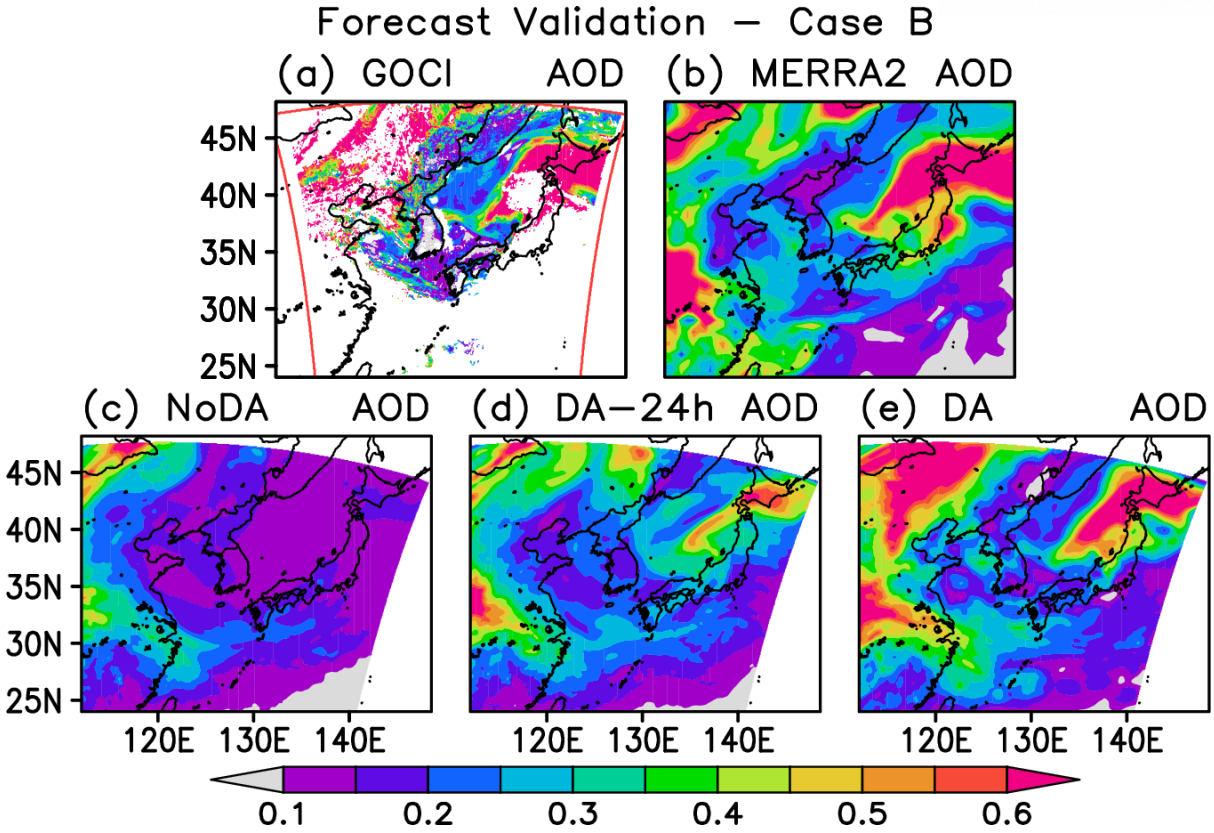


Figure 3-13. The AOD distribution at 00 UTC on 20 May 2016 over Northeast Asia from (a) the GOCI observations, (b) NoDA (with no assimilated observations), (c) DA1 (the assimilation of GOCI only), (d) DA2 (ground PM only), and (e) DA3 (GOCI and ground PM). The area within the red frame in (a) indicates the spatial coverage by the GOCI instrument. The unit of AOD is dimensionless.

For the long-range transport case, NoDA already shows the transporting pattern connected from China to Korea. It shows high AOD and is connected to Japan. The 24-hour forecast simulates a thicker AOD transporting pattern like MERRA2. The interesting part is comparing Fig. 3-14d and 3-14e, in which the model forecast from previous data assimilation overestimates the AOD over the inland of China. This might be the problem which the model resolved overestimation of China region. However, although the model forecast might produce overestimation, the overall forecast quality of 3 different cases of KORUS-AQ showed significantly improved pattern and concentration of aerosol compared to the background model. Similar to Fig. 3-12, the transporting patterns become synoptic patterns from the previous simulation, The NoDA, and 24-hour simulations show relatively quite similar to each other, more than the case A. Comparing the 24-hour simulation and DA case, the major transporting pattern is also quite similar to each other except for inland of China. This long-range transportation is continued for a long time, that the pattern itself dominated by the static meteorological condition and the major transporting direction and velocity might not be changed much for the 24-hour forecast. A different characteristic between cases A and C is the actual AOD in the transporting pattern is similar between NoDA and 24-hour forecast. This transportation is less affected by the data assimilation and mostly dominated by the model simulation.

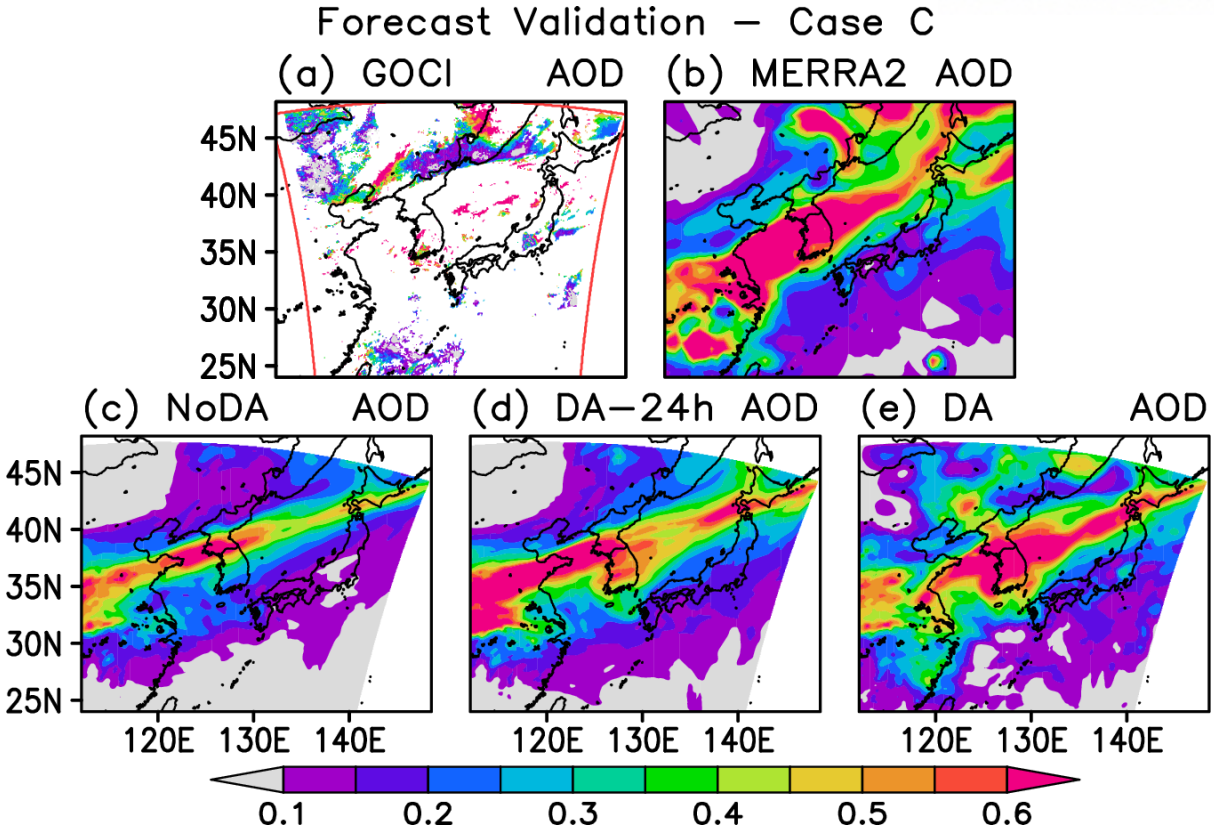


Figure 3-14. The AOD distribution at 00 UTC on 26 May 2016 over Northeast Asia from (a) the GOCI observations, (b) NoDA (with no assimilated observations), (c) DA1 (the assimilation of GOCI only), (d) DA2 (ground PM only), and (e) DA3 (GOCI and ground PM). The area within the red frame in (a) indicates the spatial coverage by the GOCI instrument. The unit of AOD is dimensionless.

3.2.2 Forecast Analysis for All-Season

Figure 3-15 shows the results of comparing the autocorrelation and autocorrelation by dividing the graph in Figure 3-15 by the seasonal correlation of the Korean Peninsula aerosol data assimilation forecast. Overall, it can be seen that the correlation is maintained above 0.5 during the 24-hour forecast in all seasons except for the April and October forecasts of PM_{10}. In October, when it exceeds 18 hours, it becomes similar to autocorrelation, but in April, it can be seen that the correlation falls below 0.5 from 7 hours. In addition, in April of PM_{10} and $PM_{2.5}$, the initial correlation starts to be lower than that of other seasons. In April, NoDA also maintains a very low correlation than other seasons. It can be said that the correlation was calculated lower in the process of calculating the balance of the error. In April, there is a yellow sand phenomenon, where sand dust is over-simulated, and the correlation of NoDA may also be low compared to other seasons. This pattern appears in the increment before and after the data assimilation and was not shown in the thesis, but as a result of comparing the dust increment of the four seasons, the only negative increment appeared only in April, suggesting that the overestimated sand dust was simulated toward reducing the data assimilation. . For this reason, it can be said that sand dust in spring is overestimated as a whole, and for this reason, the correlation may appear lower than in other seasons. The data assimilation forecast showed better performance than autocorrelation in January and July, similar to autocorrelation in October, and slightly inferior in April. The peculiar thing is that even in NoDA, the correlation increases as the forecast time passes, which is thought to be the uncertainty of the model itself. The correlation between DA and NoDA generally becomes similar after 24 hours, showing that the uncertainty of the model itself is greater than the data assimilation effect. Including the shading information, the 95% confidence level of the forecast variations for both forecasts is overlapped after the +6 hours of the forecast. It means that the absolute forecast difference is maintained until the 6 hours has been simulated. After the 6 hours, the forecasts might be similar to each other, or even worse, the DA forecast becomes inferior to the NoDA forecast. The average forecast of DA is maintaining a correlation of over 0.5 for 24 hours, but absolute reliability is maintained for 6 hours compared to the NoDA forecast.

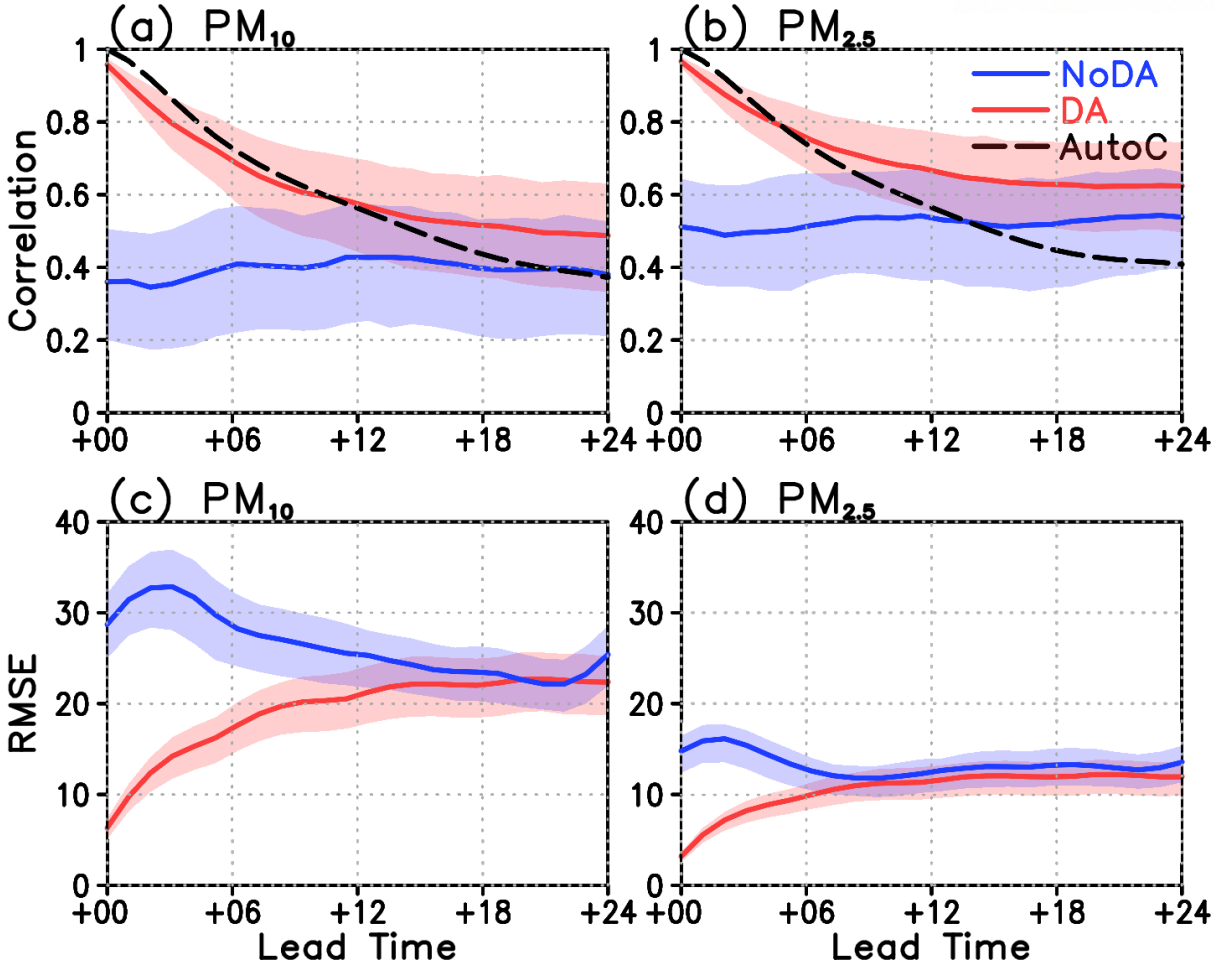


Figure 3-15. Correlation and RMSE graph for 24-hour forecasts between NoDA and DA. The correlation graph includes the persistence curve from NAMIS observation. The Black dashed line is the persistence curve of NAMIS observation. Unit of RMSE is ug m^{-3}. Shaded regions are a 95% confidence level of forecast accuracy for each simulation.

In Figure 3-16, the data assimilation effect in the forecast adapting the assimilated initial conditions shows a pattern that rapidly decreases in spring and autumn. This pattern seems to be highly related to the seasonal difference in the average correlation of the NoDA experiment. The method of the variational data assimilation is finding the balanced analysis between model and observation, and when the difference between model and observation is increased, the analysis value computed by data assimilation also might be degraded more from the original observation value. Therefore, to increase the accuracy of data assimilation, it means that a certain degree of accuracy must be guaranteed in the model as well as observation. When contrasting this with the results of Figure 3-16, the fact that the forecast performance of the model changes significantly depending on the season means that the model systematic error is large enough to degrade the forecasts between seasons or the accuracy of the input data is highly likely to be poor between the seasons. In the case of the systematic error of the model itself, the variation over time during model forecasting and the very high degree of disorder of the scatter plot shown in NoDA in Figure 3-8 shows that the structural error of the model can be large. The RMSE shows another information compared to correlation, that the mean RMSE is higher in January and April than July and October. Unlike correlation, from the perspective of RMSE, July and October show more reliable forecasts than January and April. However, the forecast skill is also revealed in RMSE results that January result shows the RMSE kept lower amount than NoDA for all 24 hours, on the other hand, results of October showed the overlapping and crossing after 18-hour forecast. Similar to Figure 3-15, shading information show 95%-confidence forecast variation for each season. Most of the absolute forecast reliability is maintained for 6 hours. The longest maintaining time is the October PM_{10} case, which is 15 hours in duration. However, this maintenance is not much reliable to the $PM_{2.5}$ case, the adequately reliable simulation becomes the July case. January and April always show the 6 or lower hour duration, which means this maintenance is closely related to the polluted state. Generally, a polluted state is affected more by the variation of the concentration. If the environment is in a clean state, the varying range of the aerosol concentration becomes lower than the high concentration.

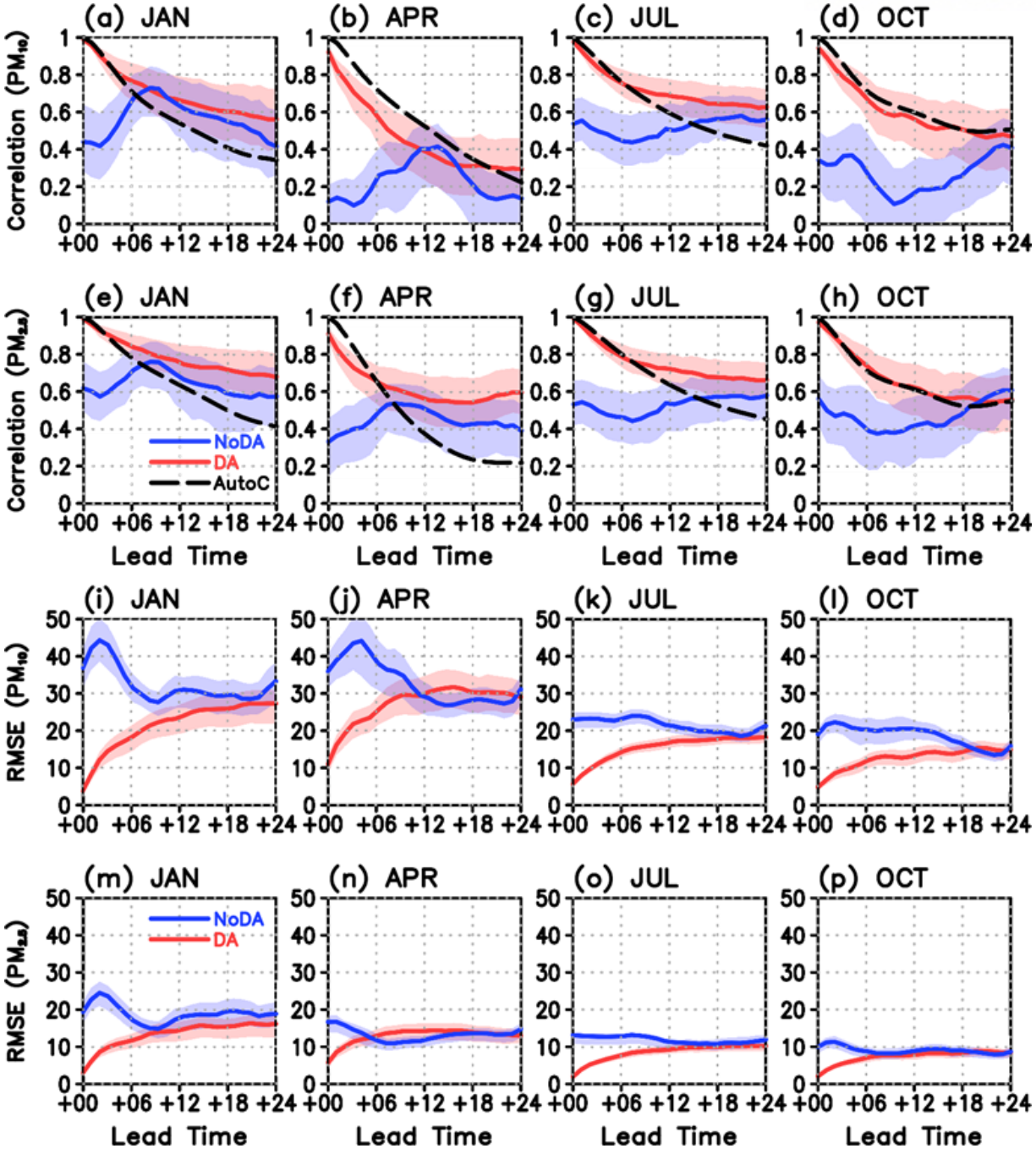


Figure 3-16. Seasonal correlation graphs between persistence curve, NoDA, and DA forecast for 24-hour forecasts. The top 8 figures (a-h) are showing correlation and the bottom 8 figures (i-p) are showing RMSE. Figures of (a, b, c, d, i, j, k, l) are showing PM_{10} and the other figures (e, f, g, h, m, n, o, p) are showing $PM_{2.5}$. Shaded regions are a 95% confidence level of forecast accuracy for each simulation.

In general, the relevance of the forecasts right after data assimilation is high because that the assimilated field still affects the simulation, but most of the operational air quality forecasts are long-term forecasts including the next-day forecast. To analyze and find the improved statistics of the next-day forecast, the entire forecast after 24 hours was summarized and compared with observations. First, the concentration distributions between observation and models are shown in Figure 3-17. When comparing the concentration distribution of NoDA and DA, the NoDA distribution is skewed toward low concentrations. It seems that the background model is not able to simulate the high-concentration cases based on the current environments. In DA results, overall concentrations are significantly increased and the PM distributions were less skewed than NoDA, but it was still not perfectly matched with the observed distribution. This imperfect match has come from the inappropriate model simulation of the fine dust concentrations higher than 80 μg m^{-3} of PM_{10} and 60 μg m^{-3} of $PM_{2.5}$. To correct the DA distributions more to observation, new adjusted values were calculated using the bias correction equation introduced in Section 2.4.2, and this is called a bias-corrected (BC) experiment.

The mean and standard deviation of the original model are removed, and values are readjusted using the mean and standard deviation calculated from observations, and the distribution of the data assimilation forecast calculated based on this is more similar to the observation, even relevant than DA. In particular, it can be seen that the high-concentration forecast is appropriately corrected, and its performance is better than the data animation forecast result before correction. From this result, to maintain the data assimilation impact longer than the 24-hour forecast, adapting the observational statistical data to the assimilated forecast is additionally required.

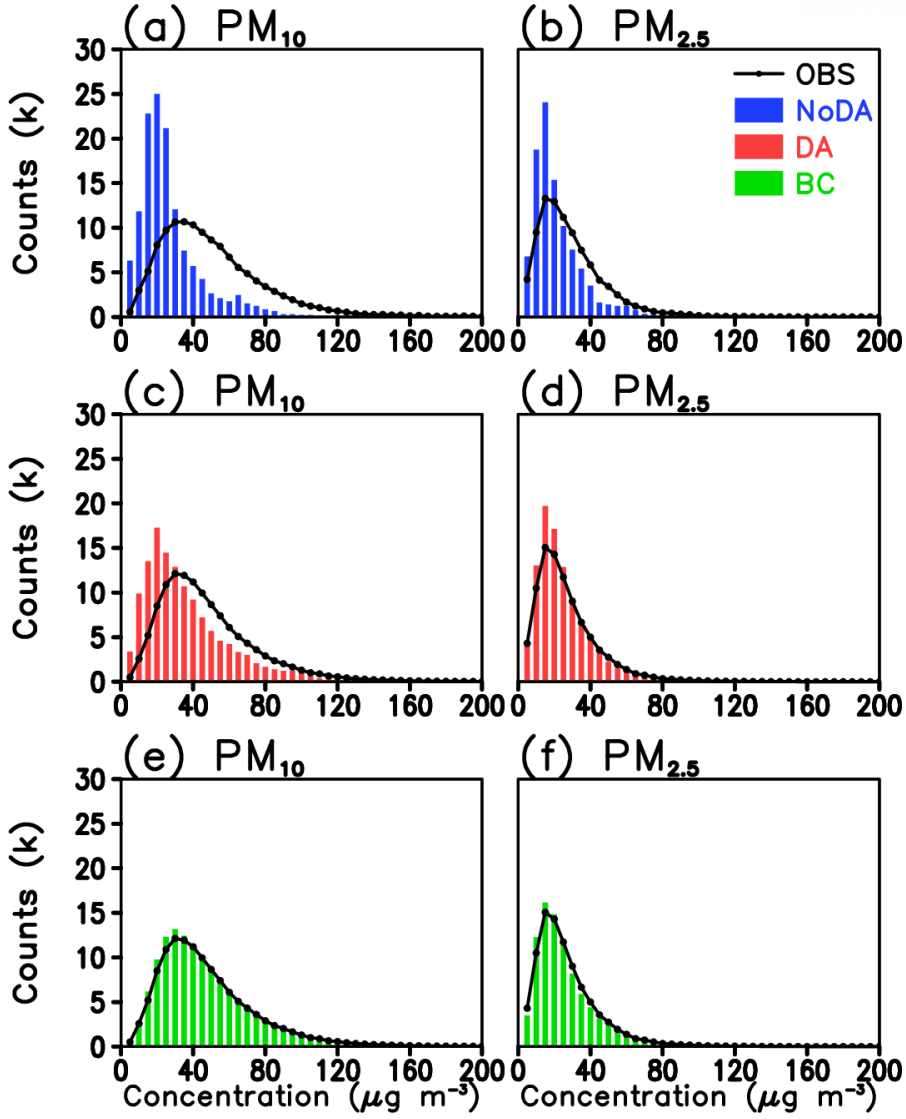


Figure 3-17. Histogram of PM concentration between observation and NoDA (a, b), DA (c, d), and DA-Bias (d, f) results of 24-hour forecasts. The left figures are PM_{10}, and the right figures are $PM_{2.5}$ results.

The results of Figure 3-17 are categorized into grade forecasts described in chapter 2, and the results of the data assimilation forecast using the data assimilation and bias correction for the hit rate, false alarm rate, and HSS are shown in Figure 3-18. In the case of NoDA, the mean hit rate was very low, within 50%. For the polluted case, it was worse than the mean case, which showed a very low hit rate within 10%, and showed a pattern that changes irregularly with time. In the case of the DA experiment, the mean hit rate decreased as the forecast duration has been passed, and the false alarm rate has been gradually increased. The DA experiment showed significantly improved forecast performance compared to NoDA due to the improved initial conditions through the data assimilation. In the initial 6-hour forecast, PM_{10} showed a hit rate of 64%, and $PM_{2.5}$ showed a hit rate of 57% and these are significantly higher than NoDA. However, the polluted case hitting rate of the DA experiment was improved than the results of the NoDA experiment, but the PM_{10} still showed a poor predicted hit rate, which is less than 40%. Comparing the bias-corrected results, the polluted case hit rate increased to 46% for PM_{10} and 54% for $PM_{2.5}$ in the 6-hour forecast, and the overall hit rate increased by about 5-10% compared to the DA experiment. In addition, as the forecast progressed, the amount of rate decrease in both the mean and polluted case was lower than the DA, and the improvement effect was maintained longer, and the 24-hour forecast was 13% higher for the polluted case than for DA. In the case of the false alarm rate, both PM_{10} and $PM_{2.5}$ were higher for the bias-corrected experiment. As shown in Fig 3-17, the false alarm rate of bias-corrected for the polluted case has been increased because the low-concentration skewness has been corrected. For the additional forecast skill analysis, the HSS skill score has been calculated. This skill score is described in section 2.4.2. In the case of the NoDA forecast, it was about 0.05, which was almost identical to the random reference forecast, and if data assimilation was performed, PM_{10} increased to 0.37 and $PM_{2.5}$ increased to 0.24. (Figure 3-18d). For the bias-corrected forecast, the score was higher than DA by 0.1. Since it is important to analyze the polluted case in fine dust forecasting, we also verified how well the polluted cases are predicted by categorizing the PM concentrations by 80 ug for PM_{10} and 35 ug for $PM_{2.5}$. (Figure 3-18e). In the case of PM_{10}, the predictive skill of high concentration cases was hardly found in the NoDA experiment, but in the 24-hour forecast of DA, PM_{10} and $PM_{2.5}$ were significantly improved from zero to 0.24 and 0.37, respectively, and after bias correction, the predictive skill improved to 0.32 and 0.41. In the case of $PM_{2.5}$, the increase in the forecast skill score due to data assimilation was very large, and the effect of the bias correction was relatively small but similar to the hit rate results, the amount of decreasing skill score followed by the forecast duration was reduced. When comparing this, it was confirmed that the data assimilation effect is well maintained until the next day's forecast as a whole. If a statistical post-processing method is applied, the forecast accuracy of about 10% on average can be further improved, and the data assimilation effect is also a longer forecast time.

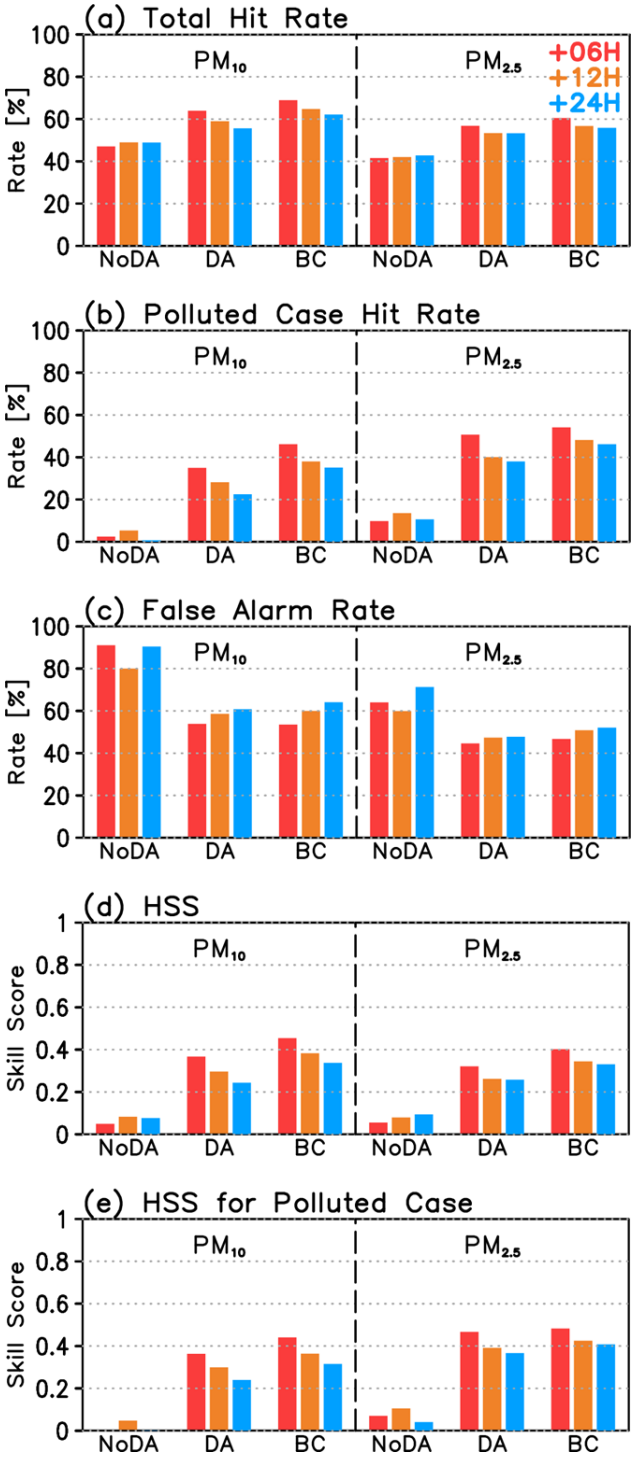


Figure 3-18. Categorical forecast statistical results of NoDA, DA, and DA-Bias for PM_{10} and $PM_{2.5}$. From top to bottom, each is showing total accuracy, polluted-case accuracy, false alarm rate, HSS, and HSS for Polluted Case. Each bar shows forecast lead time, and left part shows PM_{10}, and the right part shows $PM_{2.5}$.

4. Chapter 4

Impacts of the Data Assimilation Components

4.1. Data Assimilation Method

The first sensitivity test has been done for the data assimilation method. Data assimilation with EnKF has been tested and compared to the 3DVAR. Figure 4-1 shows the 10-day timescale graph for PM_{10} and $PM_{2.5}$ over South Korea. 3DVAR result shows slight overestimation for certain points, and EnKF shows underestimation. In PM_{10}, 3DVAR shows overestimation, however in $PM_{2.5}$, it seems that EnKF shows underestimation and shows slightly higher bias than 3DVAR. From the temporal perspective, EnKF seems to be more reliable than 3DVAR. For the EnKF, comparing EnKF_M and EnKF_N, EnKF_N shows slightly more reliability than EnKF_M. For the ensemble perspective, initializing the meteorological state for every 00UTC is considered as not only organizing the biased meteorology by model systematic bias to the original input but also reduces the current perturbation to zero points. Atmospheric transport and convection are one of the important aspects of determining atmospheric aerosol concentration, reducing the meteorological perturbation are affecting the data assimilation quality of the EnKF.

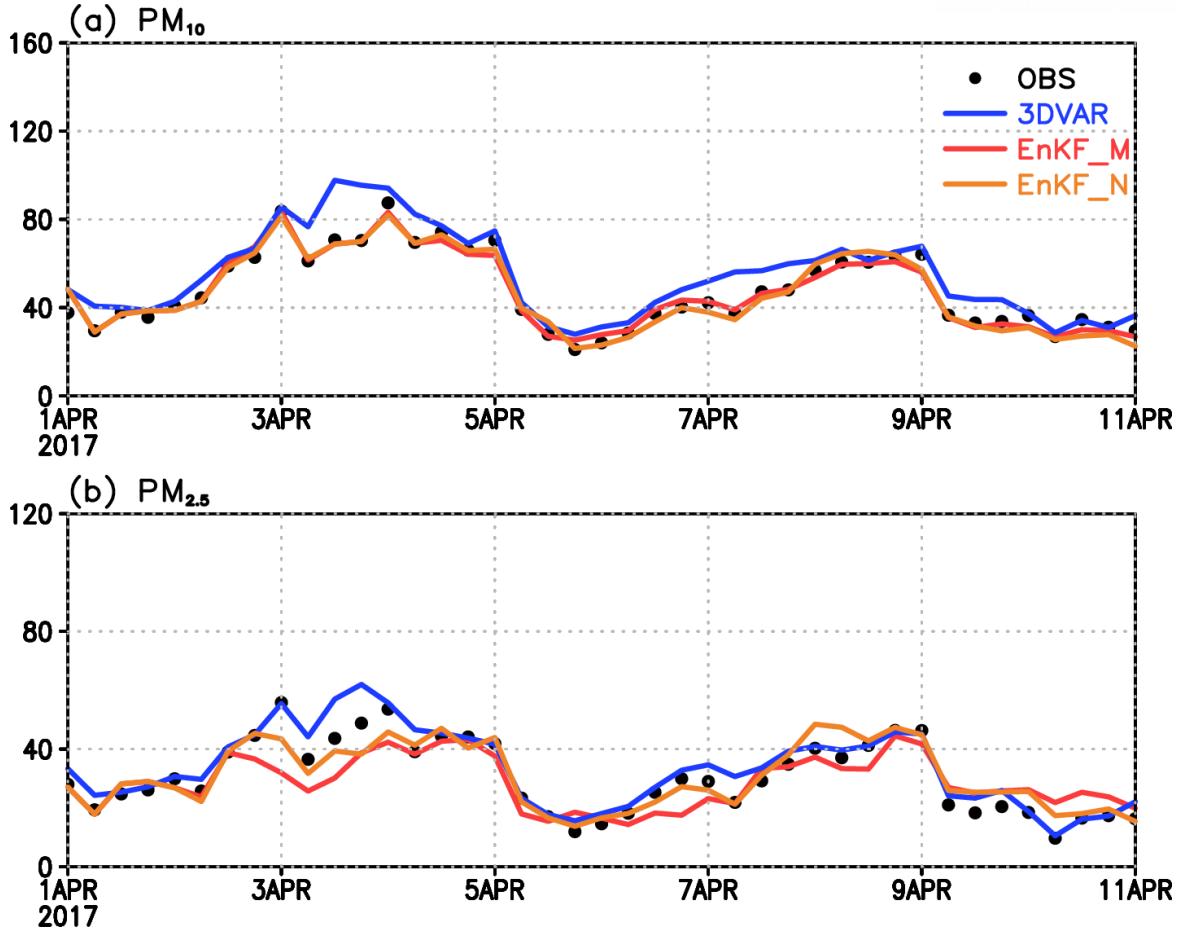


Figure 4-1. The timescale graph of the observation, 3DVAR, EnKF_M, and EnKF_N results. Each figure is shows (a) PM_{10} and (b) $PM_{2.5}$. Each data are collected from NAMIS stations, and 6-hours interval same time as data assimilation time.

Figure 4-2 shows the data assimilation quality between NAMIS and model PMs. Overall correlation showed that the 3DVAR results are the best. Between EnKF results, EnKF_M and EnKF_N, EnKF_N result showed the better organization than EnKF_M. In $PM_{2.5}$, the EnKF_N showed a significantly increasing correlation and reducing RMSE than EnKF_M. Related to Figure 4-1, maintaining the perturbation for ensemble increases the EnKF performance. However, neither of the EnKF results are showing better performance than 3DVAR, in which the 3DVAR produces better PM analysis data, at least for the assimilating period. The characteristics of the EnKF results are that mean bias is very small, compared to the 3DVAR. 3DVAR shows overestimation, but EnKF results separated more even. Because of that, the RMSE for EnKF is become smaller than the 3DVAR results, because of the overestimation of 3DVAR. However, for the current state, the EnKF method might be comparable to 3DVAR, but it cannot surpass the 3DVAR's data assimilation skill.

Comparing EnKF_M and EnKF_N, the reliability of the surface PM data has been significantly increased. The major difference between the two experiments is the meteorological initializing process. This process is generally done to prohibit the model deterioration from the model systematic errors and inaccuracy but from the perspective of the ensembles, this initialization process also resets and reduces the perturbation between ensembles, which means that the spread and RMSE relationship of two experiments should be compared to figure out how much variables have been changed.

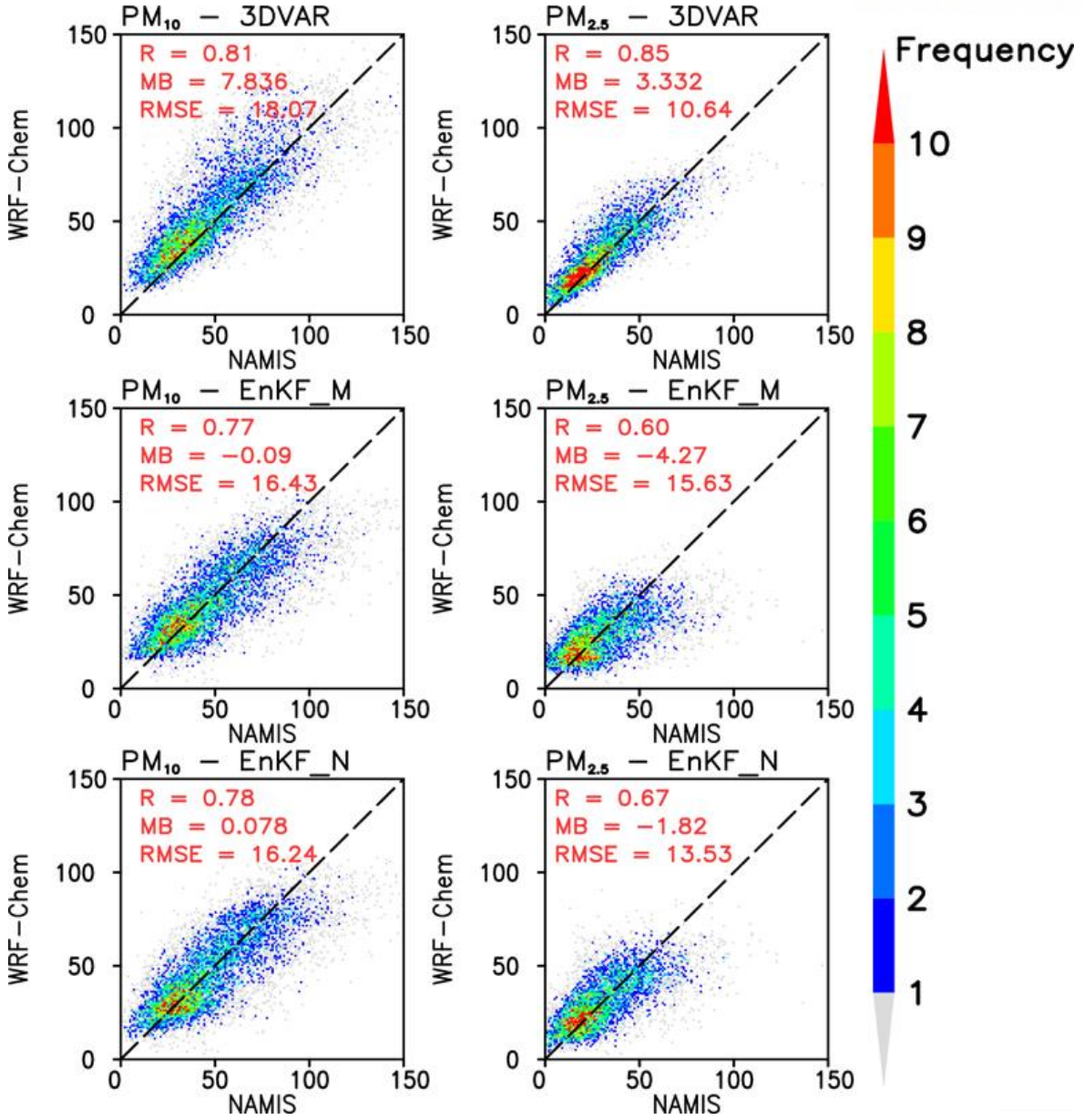


Figure 4-2. Frequency scatter plots between NAMIS observations (x-axis) and the model analysis (y-axis) for (a, c, e) PM_{10}, and (b, d, f) $PM_{2.5}$. The model analysis in (a, b) are 3DVAR, (c, d) are EnKF_M, and (e, f) are EnKF_N. Frequency indicates the number of samples in each corresponding bin. The values at 00, 06, 12, and 18 UTC values are used.

To investigate the relationship between RMSE and the spread of the EnKF method, a scatter plot between RMSE and the spread has been shown in Figure 4-3. In EnKF, RMSE shows the overall analysis bias to the observation, and the spread means the standard deviation of the variables of the ensemble. Optimal EnKF performance can be achieved when the RMSE and the spread values become similar to each other. However, in Figure 4-3, the results are showing the RMSE of EnKF means is way higher than the spreads. Small spread means that the EnKF model cannot maintain its perturbations properly, which is the crucial part to generate the background error covariance. Both of the EnKF experiments are showing a smaller spread than RMSE, which means the EnKF is not in its optimal state. However, in the EnKF experiments, comparing EnKF_M and EnKF_N shows interesting differences in the relationships. EnKF_N shows smaller RMSE, and a higher spread value than the EnKF_M results. It means when the original state shows a small spread and high RMSE, and when the spread is increased by maintaining its perturbation, then not only the spread is increased but also RMSE becomes decreased. The optimal state is RMSE equals to spread, which means that the RMSE and spreads have a positive relationship and the correlation will become 1 when the EnKF becomes an optimal state.

From the relationship of the RMSE and the spread, EnKF is hard to generate high-quality data assimilation unless the ensemble spread problem has been solved.

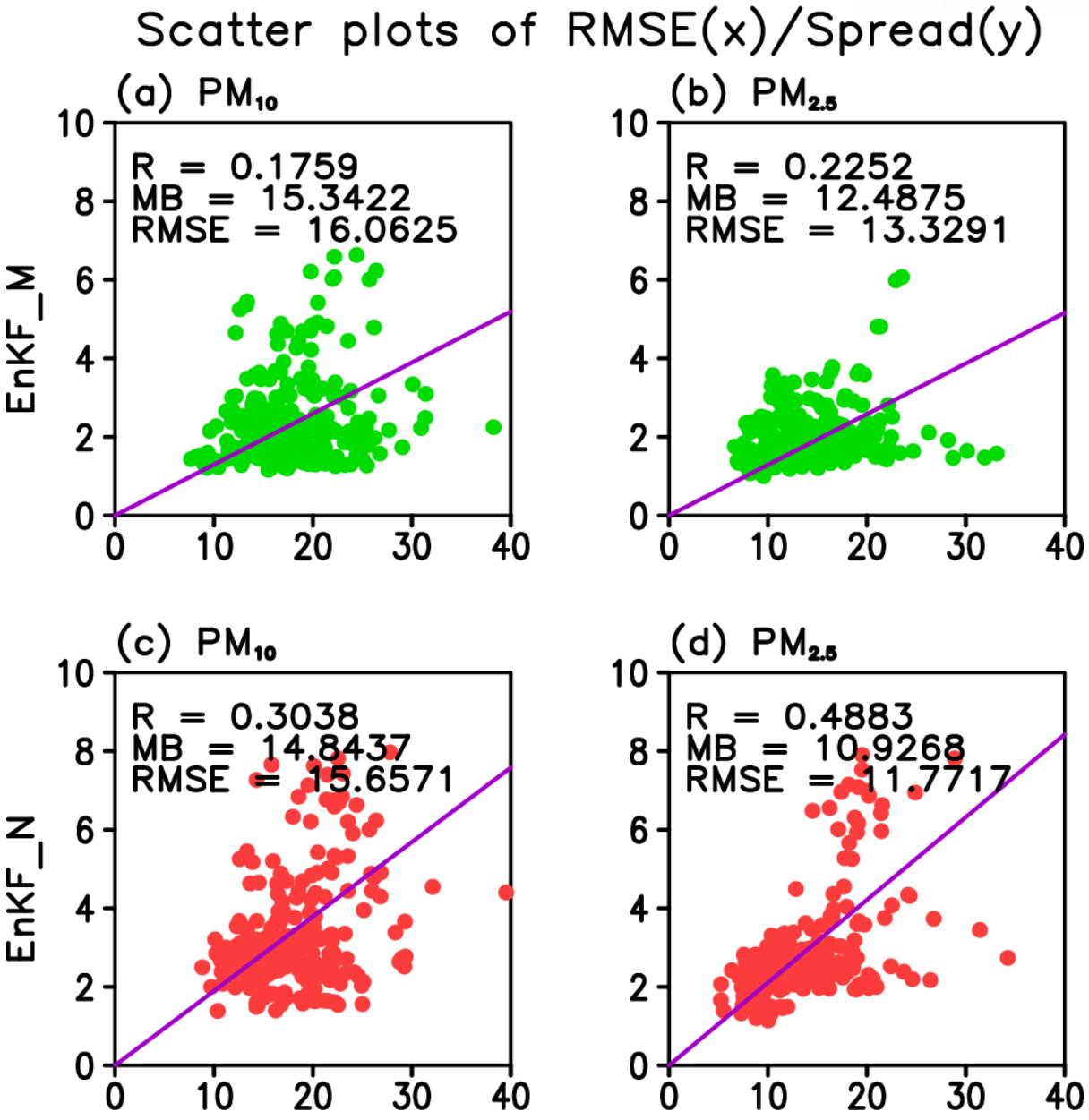


Figure 4-3. Scatter plots between RMSE of mean EnKF analysis (x-axis) and spreads of EnKF ensembles (y-axis). (a, b) are results of EnKF_M, and (c, d) are results of EnKF_N. (a, c) are PM_{10}, and (b, d) are $PM_{2.5}$ results.

4.2. Observing System Experiment

The second sensitivity test has been done for the OSE. Data assimilation with observation denial has been tested.

Figure 4-4 compares the mean surface PM_{10} from the surface observation sites between the NoDA and three OSE experiments. The observation (Fig. 4-4a) shows high values in northern China, particularly over the Beijing-Tianjin-Hebei (BTH) area and Yangtze River delta located in Shanghai. BTH area is known for the industrial complexes of manufacturing and burning fossil fuels, and a large number of transportations and wide agricultural regions. South Korea showed a relatively lower surface PM concentration compared to China.

NoDA (Fig. 4-4b) shows an underestimation of PM_{10} concentrations for both China and South Korea. The exception is the inner land of central China, where the model simulated the PM_{10} concentration similar or slightly overestimated than the observation. A similar underestimated result is already shown in Fig. 3-2, which means that the underestimation of PM_{10} is related to the model systematic errors because of the underestimation of the entire KORUS-AQ period. There might be many reasons that the model underestimated the PM_{10} concentration, such as the inaccurate emission inventory, secondary chemical formation, meteorological transportation, or dry and wet deposition processes. To define which feature is the dominant property that occurs the underestimation of surface PM_{10}, additional sensitivity experiments might be required, but in this study, the additional sensitivity experiments will not be done and it will be assumed that the model deficiency is related to the emission inventory. The current emission inventory of the EDGAR-HTAP is produced by the bottom-up approach by integrating reported emissions. Possibly there should exist many unknown sources or unreported emissions in this area. In the meantime, the model also suffers from the natural dust transport process. The NoDA run underestimates the surface PM_{10} concentration in northwest China, where the natural dust is originated from the Gobi Desert and transported to increase the aerosol amount. The GOCART module in the WRF-Chem model enables the dust scattering process that lifts the natural dust dynamically by surface winds. The overall weak surface PM_{10} concentration level in the northwest suggests a weak transport of natural dust in springtime by the model, which is shown in Fig. 3-3.

In the DA1 result (Fig. 4-4c), the surface PM_{10} has been increased for the overall regions. In the difference from NoDA (Fig. 4-4f), the GOCI AOD assimilation occurs increase of the surface PM_{10} over the land area in northern China of Manchuria, Korean Peninsula, north of Japan, and the adjacent oceans. The results show more consistency with the surface observations over the land regions. The

AOD data assimilation is mostly affected for the wide region rather than the local scale. DA1 shows the overall increasing pattern for China, but it did not resolve the local high-concentration pattern near the BTH region. For Korea, DA1 still underestimates the PM_{10} concentration than observation.

In DA2, (Fig. 4-4d), it showed that surface PM assimilation can resolve the local-scale sources in major cities in northern China by substantially increasing the PM_{10} level comparing with NoDA. In addition, the surface PM data assimilation can affect not only the location of the surface station but also the surroundings of the station. The data assimilation impact is most significant where the surface PM observations are available, while it is smallest in North Korea and Japan, where no surface observations are provided (Fig. 4-4g). Unlike the satellite AOD data, the regional representativity of the surface station is quite narrow. The adjacent region of the surface station can be improved by the surface PM data assimilation, but the affecting range is not far from the station and it is hard to resolve the observation-vacant region. The surface PM_{10} analysis is mostly benefited by assimilating both satellite and ground observation. DA3 (Fig. 4-4e) performs the best in describing the time-averaged PM_{10} distribution. The difference from NoDA (Fig. 4-4h) shows the features of simply adding two different patterns of DA1 and DA2 from NoDA (Fig. 4-4f and 4-4h).

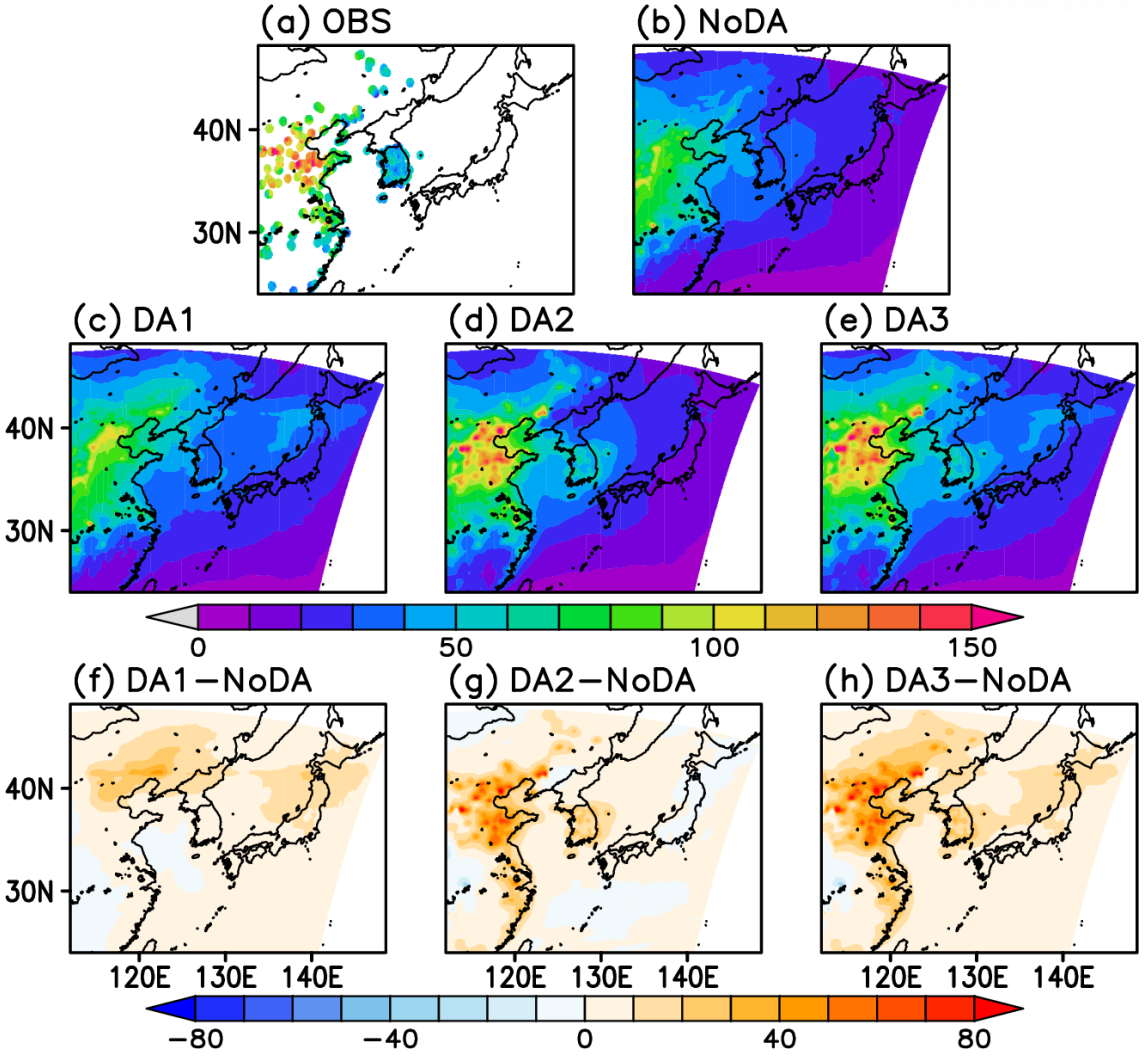


Figure 4-4. Time-averaged surface PM_{10} concentration (unit: μg m^{-3}) over Northeast Asia during KORUS-AQ from (a) surface observations, (b) NoDA, (c) DA1, (d) DA2, and (e) DA3 experiments. From (f) to (h), the figures show the differences for DA1, DA2, and DA3 from NoDA, respectively. The analysis fields at 00 and 06 UTCs are averaged.

The OSE results for $PM_{2.5}$ (Fig. 4-5) showed a higher concentration of NoDA than observation compared to the PM_{10} results. Unlike the PM_{10}, WRF-Chem can simulate the $PM_{2.5}$ and much more related to the observation. The observed (Fig. 4-5a) and the model background (Fig. 4-5b) distribution for $PM_{2.5}$ is quite similar as in PM_{10} with absolute magnitude differences only, the model simulated the PM2.5 even higher than observed data, especially over the center of China. The GOCI AOD assimilation increases the $PM_{2.5}$ even higher than NoDA (Fig. 4-5f), which is similar to the PM_{10} result. A closer look at the differences in Fig. 4-5g and 4-5h suggests the surface PM data assimilation helps suppress the overestimation bias in the continental interior by NoDA. This is not driven by the satellite data assimilation (Fig. 4-5f) as it is out of the GOCI retrieval domain. Overall, the $PM_{2.5}$ increments are not much significant as PM_{10}, but it generates the appropriate concentrations. Because of the overestimation of the PM2.5 in NoDA, the local pattern in the bias figures are not as significant as PM_{10}. DA3 can simulate both increasing patterns over the Manchuria field and decreasing patterns over the inner China region, the data assimilation using multiple observations can resolve the opposed effects and averaging the increment between single data assimilation results.

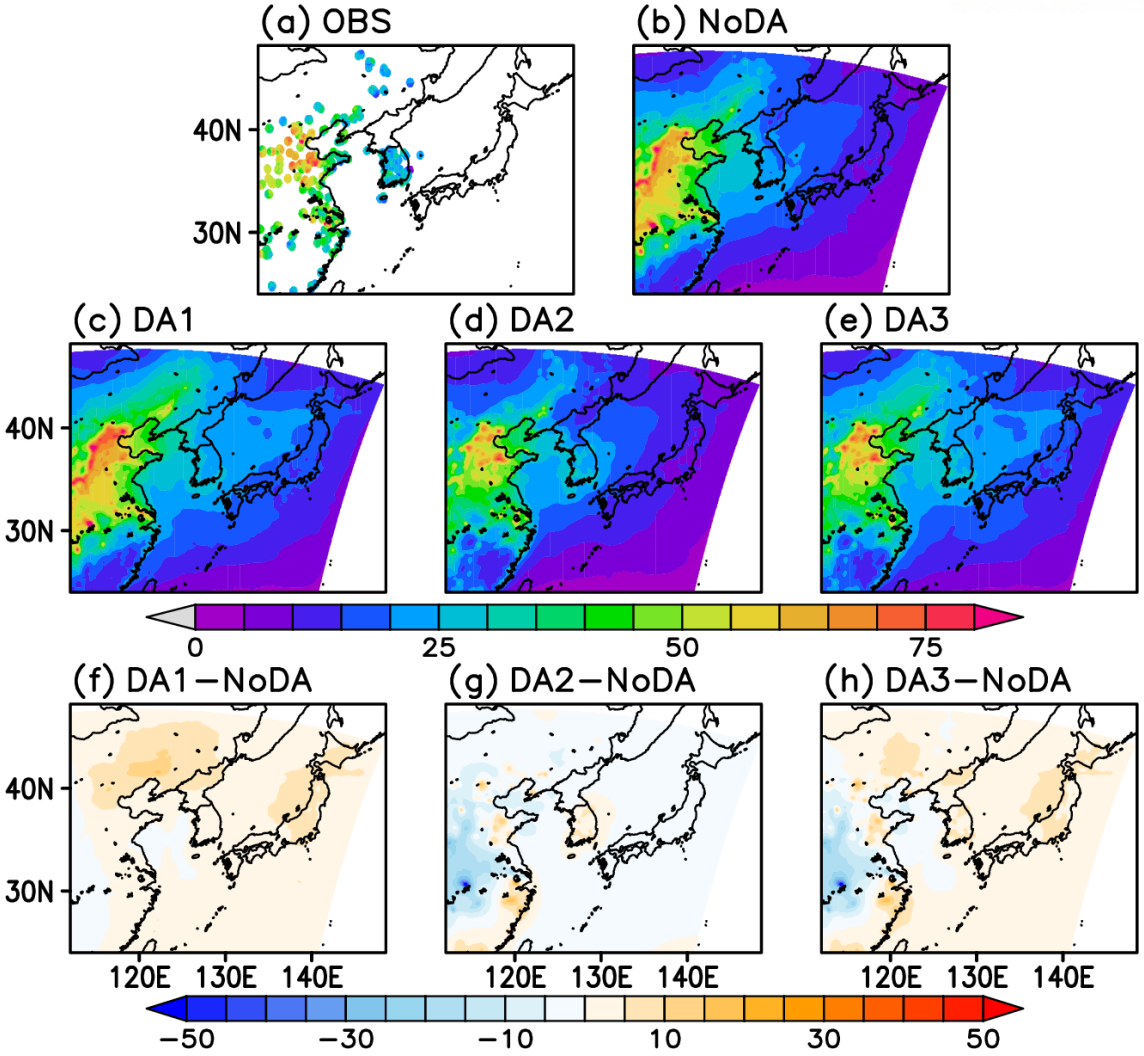


Figure 4-5. Time-averaged surface $PM_{2.5}$ concentration (unit: μg m^{-3}) over Northeast Asia during KORUS-AQ from (a) surface observations, (b) NoDA, (c) DA1, (d) DA2, and (e) DA3 experiments. From (f) to (h), the figures show the differences for DA1, DA2, and DA3 from NoDA, respectively. The analysis fields at 00 and 06 UTCs are averaged.

Figure 4-6 compares the AOD analysis by WRF-Chem. The AOD product by MERRA-2 is used as a reference from the independent analysis. The analysis data from WRF-Chem is the average of data at 00 and 06 UTC hours, only for the analysis times, respectively. The MERRA-2 reanalysis shows the large AOD values in Northeast Asia with the maximum in central China, BTH, and the Manchuria region. The DA1 (Fig. 4-6d) results in the overall increasing patterns from NoDA (Fig. 4-6c). The maximum AOD regions in the domain are also consistent with the MERRA-2 results but slightly lower amount than MERRA-2 (Fig. 4-6b). The increment is particularly large in northern China and Manchuria (Fig. 4-6f).

Because of many factors, such as the differences in the assimilated data (MODIS versus GOCI), the aerosol transport model (GEOS versus WRF-Chem), and the specified surface anthropogenic emission inventory, it is not easy to identify which factors are more responsible for the difference in the time-averaged AOD analysis. MERRA-2 shows much higher sulfate aerosols in our investigation, which might suggest the discrepancy in the anthropogenic emission between the analyses (not shown).

The DA2 (Fig. 4-6e) showed the decrease of the AOD values slightly, mostly over China, but its quantitative contribution is limited (Fig. 4-6h) and much less than the contribution by the satellite data assimilation (Fig. 4-6g), especially over the Korean Peninsula. It again suggests that the surface aerosol concentration may not relevant to the total column aerosol concentration and it is hard to resolve when most of the aerosol is remain in the upper atmosphere. The overall increase in the AOD analysis fields demonstrates the benefit of satellite data assimilation, particularly in the area of sparse observations such as in Manchuria and the data gap regions in the adjacent oceans and Japan where no surface observation is provided. When both the satellite AOD and surface PM observations are provided to the data assimilation in DA3 (Figs. 4-6f and 4-6i), the results are not much different from the case of DA1. This is more or less expectable because it suggests that the direct assimilation of the AOD variable from the satellite performs best for the AOD analysis. Comparing the AOD values in DA1 (Fig. 4-6d), DA3 (Fig. 4-6f) shows a slight decrease of the AOD values by the simultaneous assimilation of surface PM observations with the satellite data. This suggests that the data assimilation performance is not always improved and often degraded by including additional observations. Our case implies the uncertain and non-linear relationship between AOD and surface PM values.

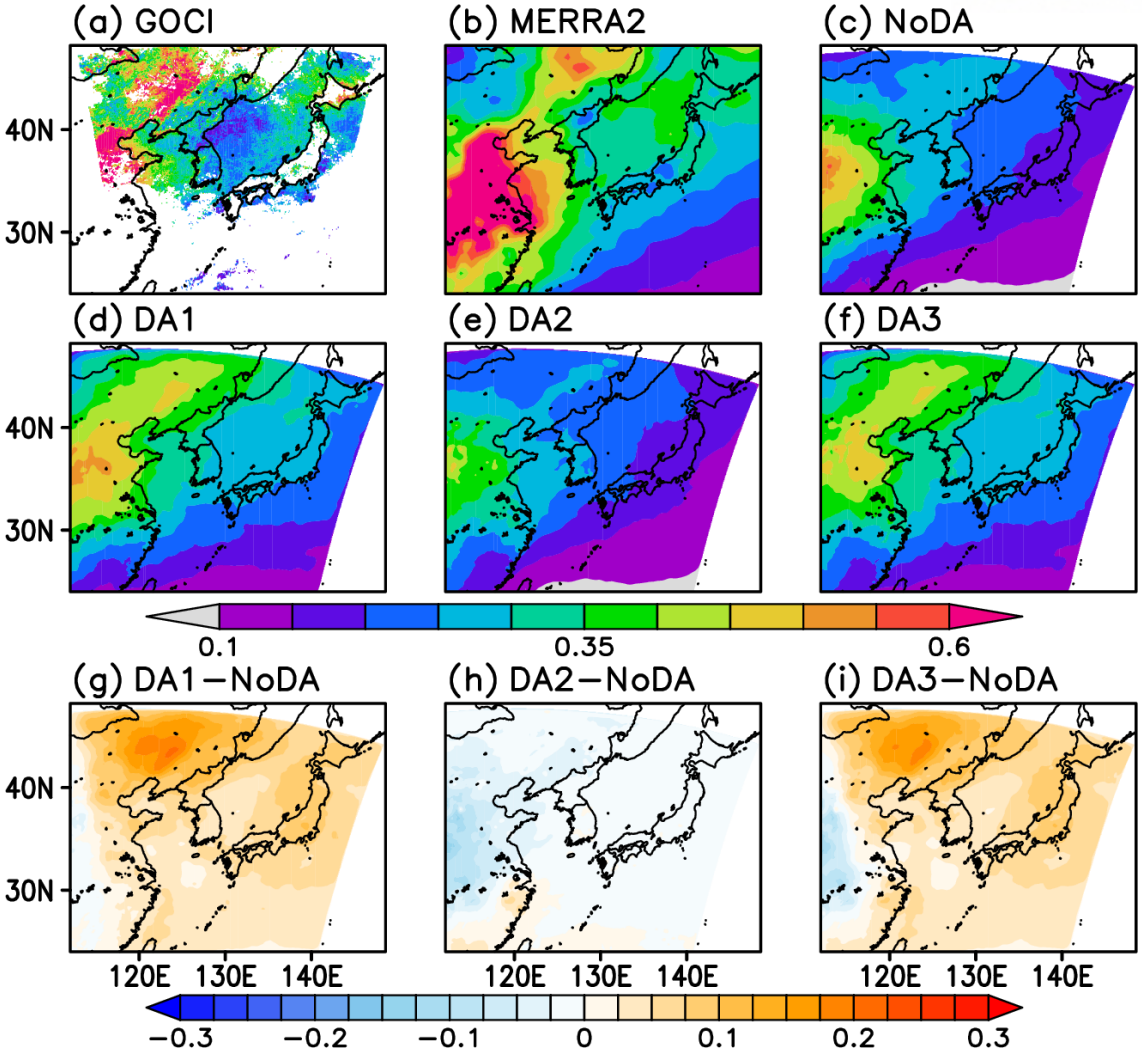


Figure 4-6. Time-averaged AOD over Northeast Asia during KORUS-AQ from (a) GOCI, (b) MERRA-2 reanalysis, (c) NoDA, (d) DA1, (e) DA2, and (f) DA3 experiments. From (g) to (i), the figures show the differences for DA1, DA2, and DA3 from NoDA, respectively. The AOD values at 00 and 06 UTCs are averaged.

Before comparing the forecast skill by different experiments, data assimilation analysis quality assessment by observations for PM_{10}, $PM_{2.5}$, and AOD, respectively, in the validation area in South Korea for the entire KORUS-AQ period has been done. Figure 4-7 shows the frequency scatter plots between observations and the model analysis from various OSEs. The correlation and the relative RMSE (rRMSE) are also shown for the quantitative metrics to measure the data assimilation performance. rRMSE has been defined by dividing RMSE by the time average of observations.

The comparison provides several details of the properties of the aerosol data assimilation. First, the model backgrounds (NoDA in Figs. 4-7a, e, and i) show unreliable results compared to the observation. All concentrations simulated without the data assimilation suffer from significant underestimations and low correlations, particularly pronounced in PM_{10} (r=0.27). PM_{10} is a more diverse concept than $PM_{2.5}$, which consists of more aerosol species and heterogeneity, PM_{10} seems to show the largest error, comparing with other variables. Second, the data assimilations provide more accurate analyses of the observed values comparing with the NoDA results. Most of the underestimated PMs and AODs are improved, and the scatter plots exhibit better alignment in the diagonal line. The data assimilation of surface PM observations provides the best fit to the observed PM values (e.g., Figs. 4-7c, d, g, and h), and the satellite AOD assimilation represents the best to the observed AOD values (Figs. 4-7j). As the AOD validation is based on the independent AERONET data, the correlation (r = 0.75) cannot be attained to such high values in PM_{10} (r= 0.90) and $PM_{2.5}$ (r=0.91). In addition, the AOD validation may be affected by the instrumental differences between top-down satellite and bottom-up ground-based AOD observations. Comparing the same type of variables between observation and model results, data assimilation showed the best quality. However, note that the satellite AOD data assimilation also improved surface PM analyses (Figs. 4-7b and f), particularly in PM_{10}. A similar improvement can be found in the AOD analysis by assimilating surface PM observations (Fig. 4-7k) vice versa. This suggests the beneficial impacts of the 3D-VAR data assimilation methods. Different types of observation data also can improve the background state, even though its effect is relatively lower than the same type of data assimilation. Finally, the data assimilation of both satellite AOD and ground PM observations (DA3) showed combined good results from DA1 and DA2. Comparing the cases for PM_{10} (cf Fig. 4-7c and d), $PM_{2.5}$ (cf Figs. 4-7g and h), and AOD (cf Figs. 4-7j and l), the changes in the correlation and RMSE seem to be negligible, or slightly degraded in terms of RMSE. This suggests that the data assimilation with multivariate observations in this 3D-VAR algorithm may degrade the consistency with the original observations. However, this is not necessarily harmful to the air quality forecast, leading to a degraded forecast skill compared with the univariate data assimilation, to be discussed in the next.

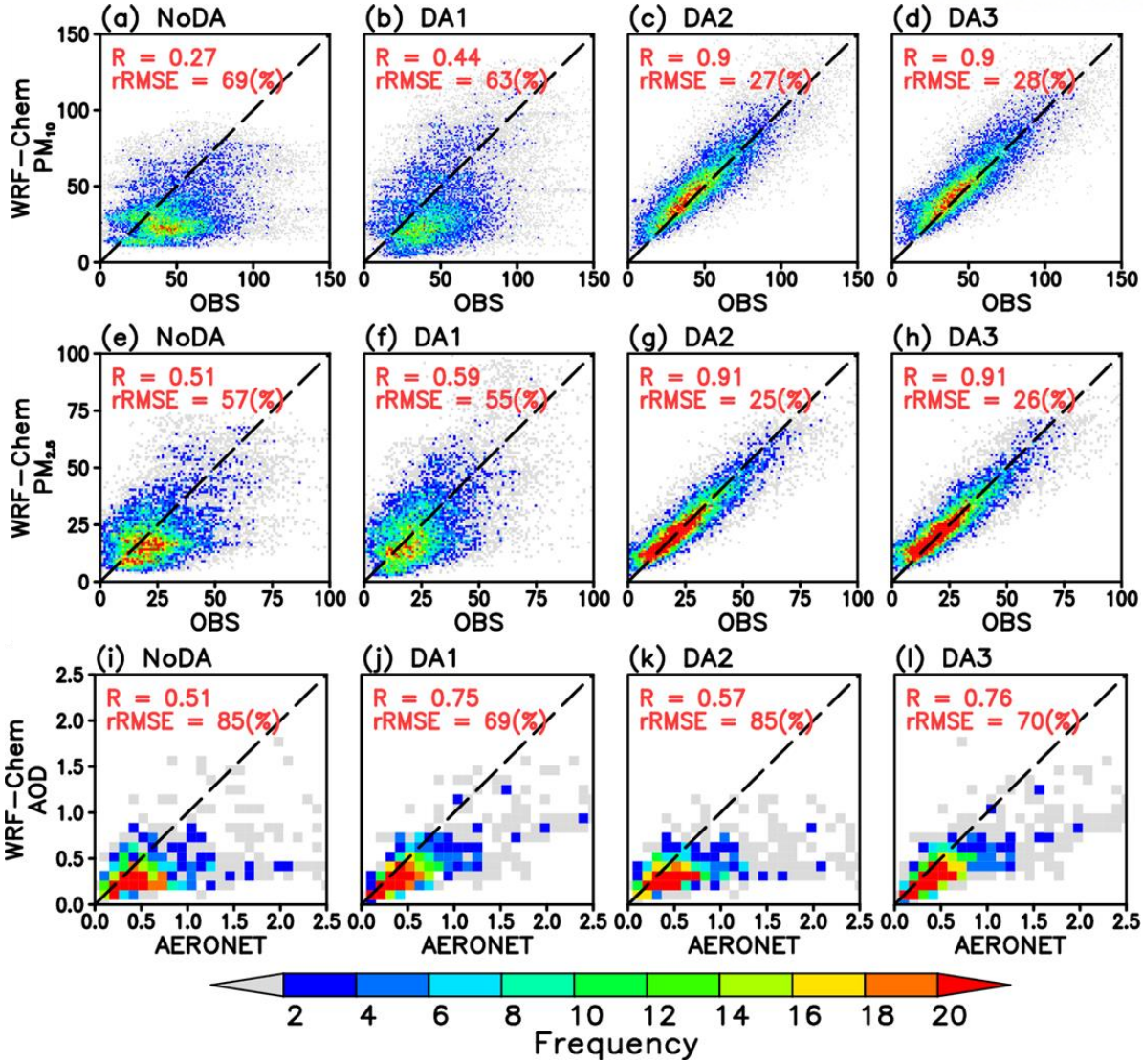


Figure 4-7. Frequency scatter plots between observations (x-axis) and the model analysis (y-axis) for (a-d) PM_{10}, (e-h) $PM_{2.5}$, and (i-l) AOD. The model analysis in (a), (e), and (i) from NoDA, (b), (f), and (j) from DA1, (c), (g), and (k) from DA2, and (d), (h), and (l) from DA3. Frequency indicates the number of samples in each corresponding bin. The values at 00 and 06 UTC values are used for surface PM concentrations and AOD. The correlation (R) and the relative root-mean-squared error (rRMSE) values are indicated in each panel. The rRMSE value is defined as RMSE divided by the time mean value of the observations.

Figure 4-8 shows the forecast skill changes followed by the forecast duration. Generally, during the forecast, the correlation gradually decreases, and the RMSE become increases in time due to the systematic biases of the model. Moreover, the forecast skill exhibits a significant difference depending on the initializing data assimilation method. The forecast of the NoDA showed the most degraded performance among the experiments. The correlation is low and below 0.3 even in the beginning. Although there is a rebound of the skill in 12 hours from the initialized, the forecast skill remains less than 0.5 during the entire forecasting period. On the other hand, the RMSE tends to decrease in time until 12 hr and starts to increase. This rebound of skill and the minimized error in about 12 hours suggests the skill recovery due to the model's capability to reproduce the observed diurnal variation of PM_{10} and $PM_{2.5}$ concentrations. For $PM_{2.5}$, this rebound is smaller than PM_{10}, and the overall forecast is more consistent and the reliability is higher than PM_{10}.

The results in Fig 4-8 indicate that the air quality forecast skill is significantly improved by the aerosol data assimilation using both satellite and surface observations. Comparing with NoDA, all of the other forecasts from the analyses of the data assimilation experiments show higher correlation and lower RMSE throughout the forecast time as long as 24 hours. This indicates that the data assimilation impact tends to last at least longer than 24 hours in this case. When the initial conditions are determined by the data assimilation with both satellite AOD and surface PM observations (DA3), the forecast skill is the best for both surface PM_{10} and $PM_{2.5}$ except for the first few hours. It maintains a correlation higher than 0.4 for all times, a useful level valid for operational forecasts. The RMSE is increasing as the forecast time increases and the data assimilation impact is gradually disappeared. Comparing with DA3, DA2 performs a considerable skill, which suggests that the data assimilation of surface PM data is one of the critical components in improving the quality of the surface PM forecasts. Note that the forecast skill of DA3 is slightly low in the beginning stages, but it overpasses the skill of DA2 a few hours after the initialization. The RMSE also becomes lower after a few hours in DA3. This implies that the initial conditions may show minor discrepancies compared with the observed, and the simultaneous data assimilation with both satellite and surface data can maintain more balanced states among the variables and improve the forecast skill in an extended time. When only the satellite data is assimilated to determine the initial states of the forecasts (DA1), the correlation skill is increasing, and the RMSE is decreasing uniformly throughout the forecast time. This is related to the less accurate calibration of surface PM concentration by the assimilation of the satellite data only (Figs. 4-8b and f). It suggests that the satellite data should have great difficulty in representing the local variation of surface PM concentration particularly in the verification regions where the surface observations exist.

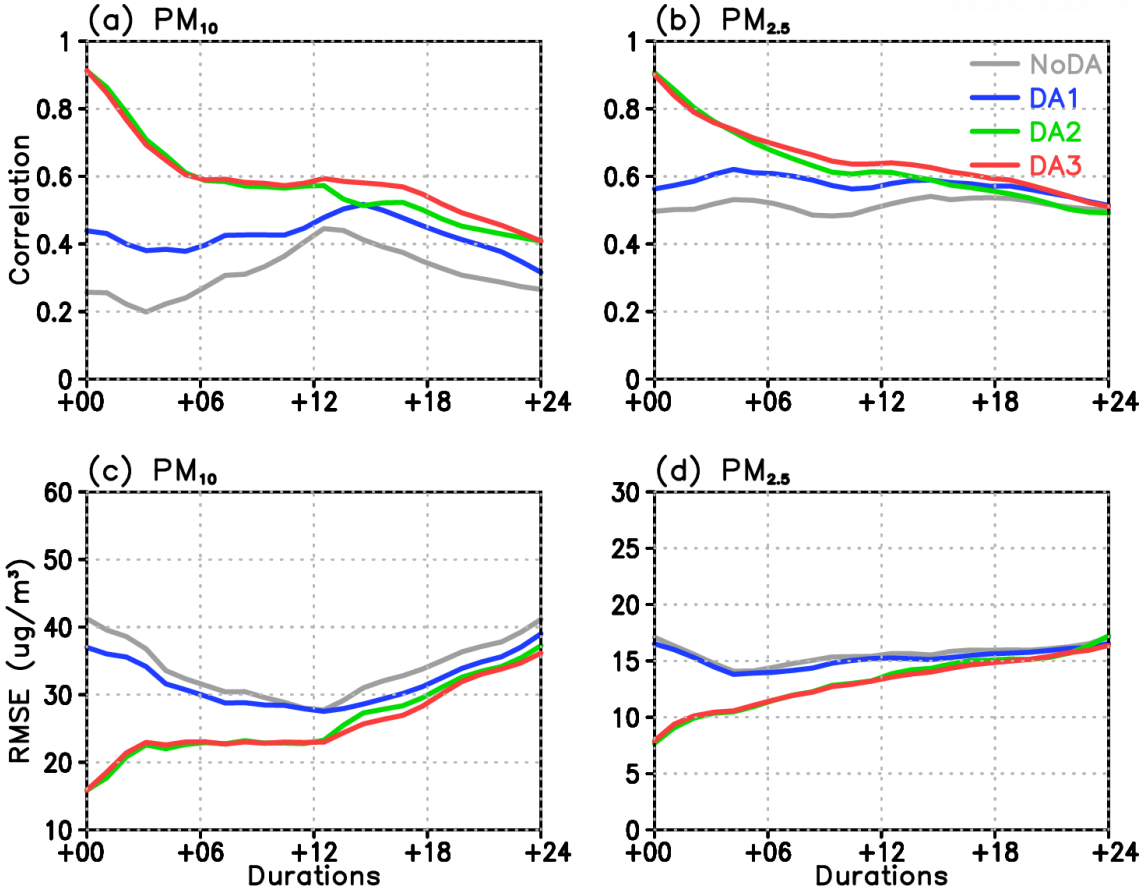


Figure 4-8. The changes in the correlation of the surface (a) PM_{10} and (b) $PM_{2.5}$ concentration forecasts and the RMSE for (c) PM_{10} and (d) $PM_{2.5}$ from the initial state until +24 hr. The forecast skill is the average over all of the NAMIS stations in South Korea from the forecasts started at 00 UTC every day for 24 hours during KORUS-AQ. Each color line indicates the cases of NoDA (grey), DA1 (blue), DA2 (green), and DA3 (red).

4.3. Constraining Meteorology

The third sensitivity test has been done for the constraining meteorological condition. Data assimilation with FDDA and without FDDA has been simulated and compared. Meteorological condition is also one of the important elements to spread out the atmospheric aerosols to the 3-dimensional grid fields. The numerical models continuously produce systematic errors and the meteorological condition is simultaneously distorted during the forecast. Data assimilation is also affected by the background model state, that the basic concept of the 3DVAR method is finding the balance between observation and model. If the model value is distorted too much, it can deteriorate the data assimilation skills under the same background error covariance matrix.

Figure 4-9 shows the PM timescale graph between FDDA and NF. FDDA (red) line shows a much similar trend to the observation than NF. Without the FDDA option, the model becomes distorted more than using FDDA, and distorted background from forecast also affects the next step of data assimilation. For each data assimilation period of 00, 06, 12, and 18UTC, the model result drops to the observation, but the drop rate and the distance from the observation become different because of the different starting points. The original FDDA only affects the meteorological condition, so the difference in the analysis period is directly affected by the previous simulation. The FDDA procedure mostly affects the u-wind, v-wind, temperature, and moisture. All of these components can affect the PM concentration. Which might be the dominant component related to the PM concentration? The next figure described the average of PM and meteorology, and bias between NF and FDDA.

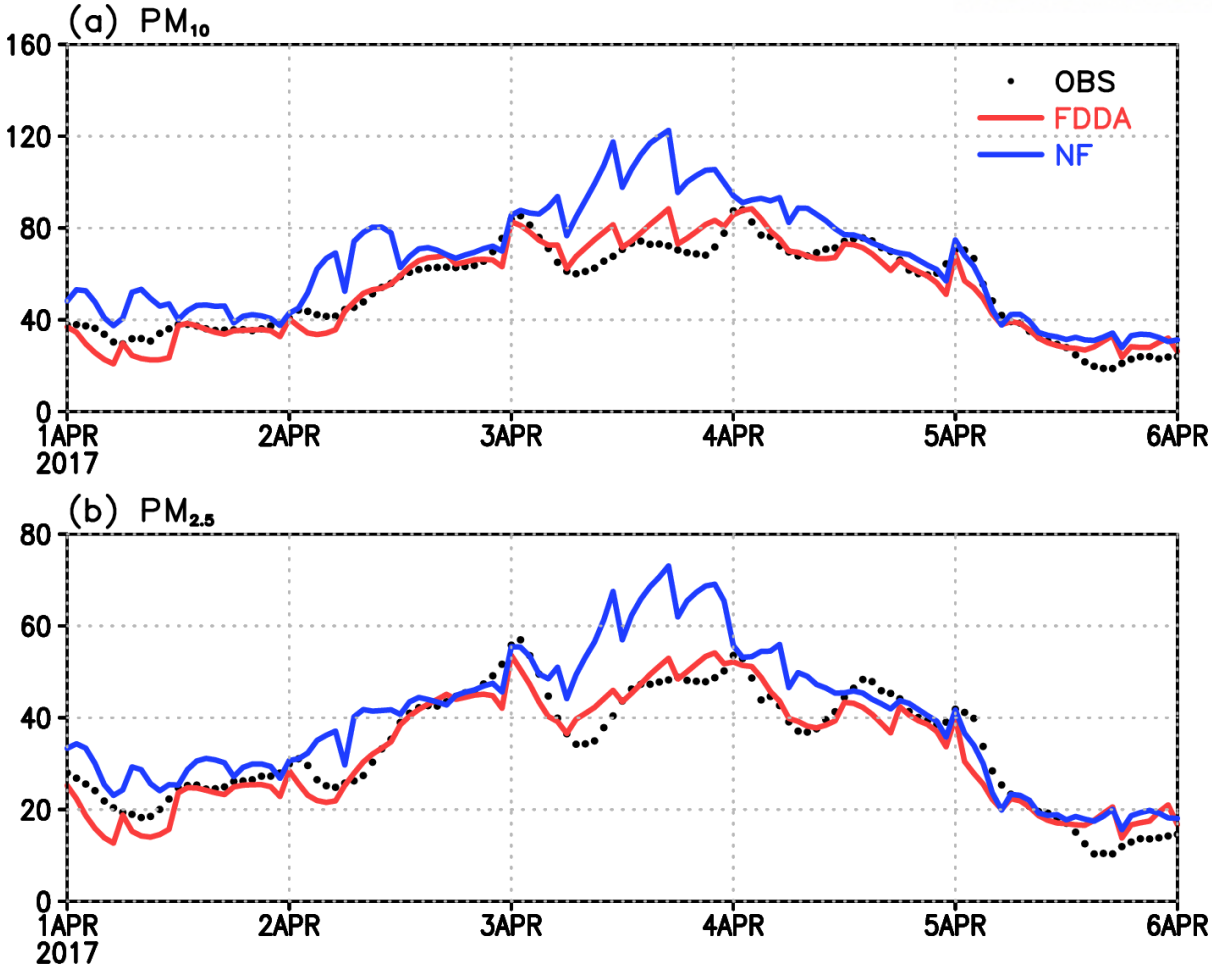


Figure 4-9. The timescale graph of the observation, FDDA, and NF results. Each figure is shows (a) PM_{10} and (b) $PM_{2.5}$. Each data are collected from NAMIS stations, and consists of a 1-hours interval.

Figure 4-10 shows the analysis of the surface PM_{10} and the surface wind speed and vectors. Comparing the surface PM_{10} concentrations between NF and FDDA, NF shows an overall high concentration than FDDA and the middle region of the Korean Peninsula shows the peak difference than surroundings. Comparing the surface wind, the NF shows slower wind speeds all over the Korean peninsula. The peak of the wind speed decrement shows in the eastern and middle part of the peninsula. The peak region of PM_{10} and wind seem different from each other except for the middle-eastern region, but including the bias wind direction, the wind direction is reversed, which means that the transporting westerly wind is much weaker in NF. This wind change is significant in the middle region of the peninsula, which is reliable to the overestimated PM regions. The wind bias vector shows that the eastern coast of the peninsula shows easterly wind bias, which the transportation to the east is more prohibited in the NF result. The wind bias of the Yellow Sea is relatively smaller than the wind bias of the East Sea, the wind flows from west to east become blocked on the land, and the aerosol can accumulate over the land, and the especially weaker wind exists on the middle region, which might be accumulated more aerosols than surroundings.

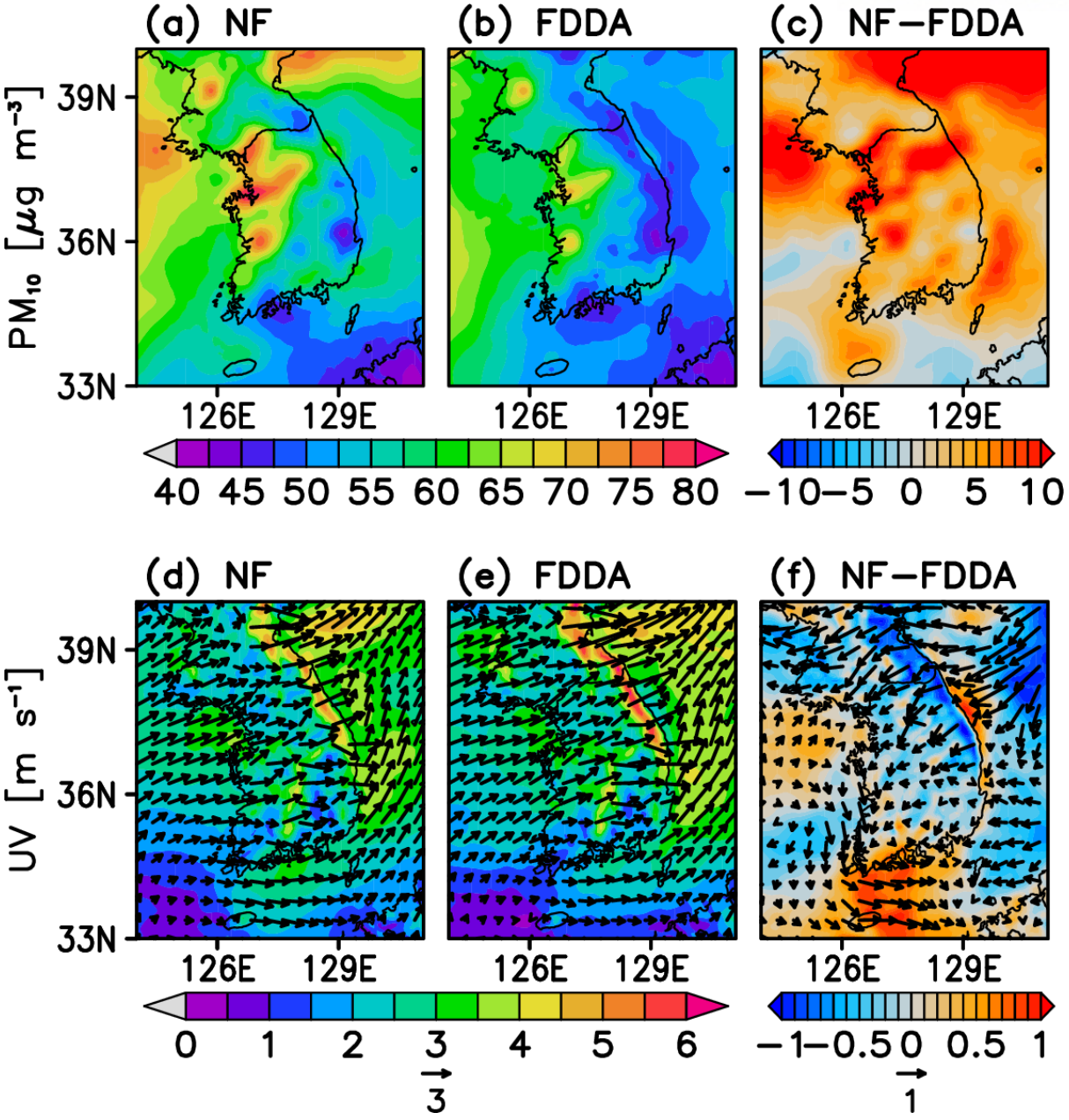


Figure 4-10. The average PM_{10} and meteorological comparisons between NF and FDDA results. (a, b, c) are showing PM_{10} concentration, and (d, e, f) are showing UV wind speeds and vectors for the surface. (a, d) are showing NF results, (b, e) are showing FDDA results, and (c, f) are showing differences between NF and FDDA.

To analyze the affection relationship between PM and meteorology, the statistics for all simulated periods might be required. Figure 4-11 shows the relationship between meteorological variables and PM_{10}. To compare the proper relationship this study assumes that the FDDA is a more reliable result than NF and the single relationship has been compared based on the FDDA result. General ideas about the relationship between PM and meteorology is described as a negative relationship for wind speed because the high wind speed flushes out the accumulated aerosols, a negative relationship for temperature because of the turbulent convection mixes the surface aerosols to the upper atmosphere, and a negative relationship for atmospheric moistures might affect both ways to reduce the aerosol concentration by wet deposition, or promotes the hydrophilic aerosols such as nitrate or sulfate. The result shows that the wind speed has a negative relationship, the temperature does not show any relationship to the aerosol, and humidity showed a negative relationship, which seems the deposition process is dominant the aerosol promotion. Comparing the bias relationship, the regression of UV wind shows a negative relationship. From figure 4-10, the negative bias of wind speed simulated higher PM_{10} concentration, the actual regression between biases are also showing the negative relationship. Temperature also shows a negative relationship, but a much weaker relationship than horizontal wind. For vapor mixing ratio, it shows a weak positive relationship, but the distribution of moisture bias seems almost even for both sides, it might be there is no relationship between moisture bias and PM. Overall, FDDA mostly affected the horizontal wind speed and directions, and for this case, the wind change made the surface aerosol concentration more reliable to the observation.

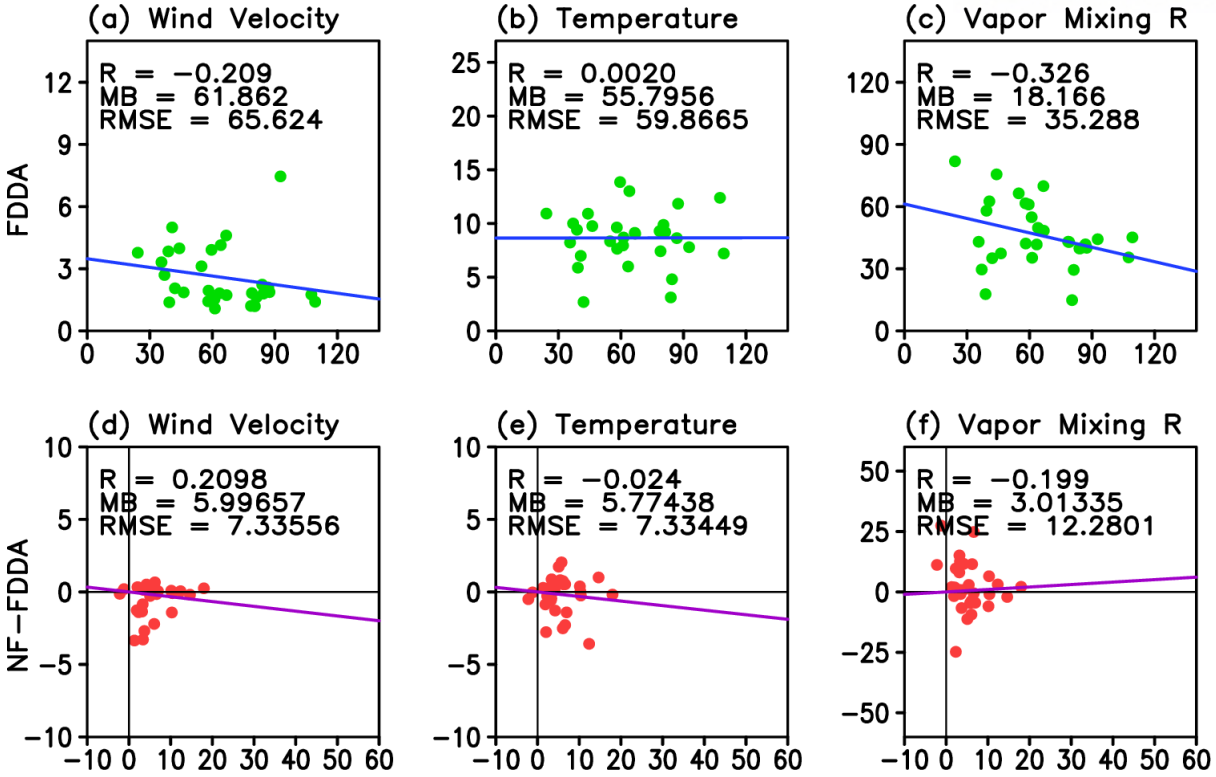


Figure 4-11. The scatter plots of FDDA and bias of NF and FDDA between surface PM_{10} and surface meteorological variables. In every scatter plot, the x-axis is defined as PM_{10} concentration. (a, b, c) are showing the FDDA results. (d, e, f) are showing the bias relationship between NF and FDDA. (a, d) are results for wind velocity, (b, e) are showing temperatures, and (c, f) are showing vapor mixing ratio.

5. Summary and Concluding Remarks

Long-term analyses provided the general data assimilation and forecast skills for Northeast Asia and South Korea. Analysis quality is significantly high which the correlation to observation is near 0.9. This improvement from the background model is enough to represent the natural statement and support further air quality forecasting. Overall forecast skills are maintained at least 6 hours for a 95% confidence level and average forecast skills are maintained at least 24 hours. Adding the bias correction method, the air quality forecast might be maintained further than 24 hours. For both KORUS-AQ and All-Season cases, the general data assimilation skill has been analyzed and proven to be reliable to the observations.

The seasonal data assimilation and forecast skills are significantly different by the seasons. The highly contaminated seasons showed fewer integration skills than clean seasons. However, the variation of the model simulation is too wide, then it also reduces the data assimilation and forecast skills. These odd cases are winter and autumn, which are higher aerosols but a smaller variation of the forecast, or clean state but highly variable forecast.

The 3DVAR aerosol data assimilation still can be adaptable, and it is comparable to the other sophisticated data assimilation methods, such as the EnKF method. For any of the data assimilation methods, processing accurate background error covariance is a crucial part. EnKF is a popular method that can process time-dependent background errors from the ensemble, however, the skill will deteriorate if the error generation has a problem such as model confidence.

Satellite and In-situ observations have their advantages, wide-range affection, and local representation. Using both data is generally stabilizing the data assimilation and forecast accuracy. Using multiple observations must be the default setup when evaluating the data assimilation system, but if the observation is limited, then an identical observation type that will be verified should be adapted.

The FDDA meteorological nudging can prohibit the model simulation process from increasing systematic bias of meteorology. If the meteorological input is reliable, the FDDA procedure significantly affects the aerosol data assimilation and forecast. Generally, the wind makes the most variances and changes the local transporting directions.

6. Possible Studies Applicable for Further Works

From chapter 4, one of the sensitivity tests was comparing data assimilation methods and analyzed the data assimilation skills between 3DVAR and EnKF. Multiple tests have been done for the sample case, but the EnKF simulation for this study has not been performed well that skill did not fit the general consideration of EnKF skills from previous studies. Proper EnKF can simulate the aerosol data assimilation significantly better than current simulation, the optimization of the EnKF system might be the proper work to process. For the study, various kind of experiments has been done to expand the actual spreads between ensembles, and the experiments such as emission perturbation, meteorological physical schemes, etc. have been affected to the expansion of the ensemble, however, not significant. For further advanced study, an optimized EnKF system should be prepared.

There are other works to approach the data assimilation method that skill is beyond the 3DVAR and EnKF method. 3DVAR and EnKF have their pros and cons of the method, and to merge the strength points of both methods has been formed as hybrid data assimilation. There are several methods to do hybrid data assimilation, and for the beginning of the research, combining 3DVAR and EnKF methods will be done as En3DVAR. The main part of the hybrid is combining the background error covariance, that using both fixed average background error and periodically generating errors from ensembles. General theories provide that the background error from an exact period can reflect a more accurate model statement than averaged error, but like the case of this study, the ensemble might not generate the error properly because of the confidence of the model and lack of perturbation of ensembles, in the case, background error from 3DVAR can support the data assimilation quality from fixed error. Generally, hybrid data assimilation overcomes the limitations of previous two assimilation method, that might generate more reliable than 3DVAR or EnKF single results. For advanced data assimilation, the establishment of a hybrid is recommended.

For the operation purpose, preparing the data assimilation base for other aerosol models is also proper work. Most of the experiment for data assimilation is focused on the GOCART aerosol model, which is used because of the proper internal modules of natural dust generation and the CRTM model to assimilate the satellite AOD data. However, there are several different aerosol models are existing such as MADE/SORGAM, MADE/VBS, MOSAIC, etc. for WRF-Chem, and looking for other models such as CMAQ, AERO5, AERO6 models are also presented. GOCART is not a proper model for nitrate formation which is also one of the main atmospheric aerosols over Northeast Asia, investigation for other aerosol models is the necessary process. Not every operating system is using GOCART models, for the compatibility to the other systems and establishing the proper data assimilation system, preparing data assimilation system for other aerosol models might be important.

References

Balkanski, Y. J., D. J. Jacob, G. M. Gardner, W. C. Graustein, and K. K. Turekian, 1993: Transport and residence times of tropospheric aerosols inferred from a global three-dimensional simulation of 210Pb. *Journal of Geophysical Research: Atmospheres*, **98,** 20573-20586.

Bellouin, N., and Coauthors, 2011: The HadGEM2 family of met office unified model climate configurations. *Geoscientific Model Development*, **4,** 723-757.

Brasseur, G., D. Hauglustaine, S. Walters, P. Rasch, J. F. Müller, C. Granier, and X. Tie, 1998: MOZART, a global chemical transport model for ozone and related chemical tracers: 1. Model description. *Journal of Geophysical Research: Atmospheres*, **103,** 28265-28289.

Carmichael, G., and Coauthors, 2008: MICS-Asia II: The model intercomparison study for Asia Phase II methodology and overview of findings. *Atmospheric Environment*, **42,** 3468-3490.

Chen, F., and Coauthors, 1996: Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research: Atmospheres*, **101,** 7251-7268.

Chin, M., P. Ginoux, O. Torres, B. Holben, D. Savoie, and J. Prospero, 1999: GOCART Model Simulated Tropospheric Aerosols and Comparison with Satellite, Sun Photometer, and Field Measurement Data.

Choi, M., and Coauthors, 2018: GOCI Yonsei aerosol retrieval version 2 products: an improved algorithm and error analysis with uncertainty estimation from 5-year validation over East Asia. *Atmospheric Measurement Techniques*, **11,** 385-408.

Choi, Y., and Coauthors, 2021: Temporal and spatial variations of aerosol optical properties over the Korean peninsula during KORUS-AQ. *Atmospheric Environment***,** 118301.

Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models.

Cimorelli, A. J., and Coauthors, 2005: AERMOD: A dispersion model for industrial source applications. Part I: General model formulation and boundary layer characterization. *Journal of applied meteorology*, **44,** 682-693.

Collins, W. D., P. J. Rasch, B. E. Eaton, B. V. Khattatov, J. F. Lamarque, and C. S. Zender, 2001: Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX. *Journal of Geophysical Research: Atmospheres*, **106,** 7313-7336.

Commerce, N. C. f. E. P. N. W. S. N. U. D. o., 2000: NCEP FNL operational model global tropospheric analyses, continuing from July 1999. *Dataset*.

Elbern, H., A. Strunk, H. Schmidt, and O. Talagrand, 2007: Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmospheric Chemistry and Physics*, **7,** 3749-3769.

Emmons, L. K., and Coauthors, 2010: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4). *Geoscientific Model Development*, **3,** 43-67.

Evensen, G., 2004: Sampling strategies and square root analysis schemes for the EnKF. *Ocean dynamics*, **54,** 539-560.

Gelaro, R., and Coauthors, 2017: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of climate*, **30,** 5419-5454.

Ghannam, K., and M. El-Fadel, 2013: Emissions characterization and regulatory compliance at an industrial complex: an integrated MM5/CALPUFF approach. *Atmospheric Environment*, **69,** 156-169.

Ginoux, P., M. Chin, I. Tegen, J. M. Prospero, B. Holben, O. Dubovik, and S. J. Lin, 2001: Sources and distributions of dust aerosols simulated with the GOCART model. *Journal of Geophysical Research: Atmospheres*, **106,** 20255-20273.

Giorgi, F., and W. L. Chameides, 1986: Rainout lifetimes of highly soluble aerosols and gases as inferred from simulations with a general circulation model. *Journal of Geophysical Research: Atmospheres*, **91,** 14367-14376.

Grell, G. A., and D. Dévényi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, **29,** 38-31-38-34.

Grell, G. A., S. E. Peckham, R. Schmitz, S. A. McKeen, G. Frost, W. C. Skamarock, and B. Eder, 2005: Fully coupled "online" chemistry within the WRF model. *Atmospheric Environment*, **39,** 6957-6975.
Grewal, M. S., and A. P. Andrews, 2008: Practical considerations. *Kalman filtering: Theory and practice using MATLAB***,** 355-426.
Guenther, A., T. Karl, P. Harley, C. Wiedinmyer, P. I. Palmer, and C. Geron, 2006: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmospheric Chemistry and Physics*, **6,** 3181-3210.
Heidke, P., 1926: Berechnung des Erfolges und der Güte der Windstärkevorhersagen im Sturmwarnungsdienst. *Geografiska Annaler*, **8,** 301-349.
Helfand, H., and J. Labraga, 1988: Design of a nonsingular level 2.5 second-order closure model for the prediction of atmospheric turbulence. *Journal of Atmospheric Sciences*, **45,** 113-132.
Holben, B. N., and Coauthors, 1998: AERONET—A federated instrument network and data archive for aerosol characterization. *Remote sensing of environment*, **66,** 1-16.
Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly weather review*, **134,** 2318-2341.
Hu, M., C. Zhou, H. Shao, D. Stark, and K. Newman, 2016: GSI.
Janssens-Maenhout, G., and Coauthors, 2012: EDGAR-HTAP: a harmonized gridded air pollution emission dataset based on national inventories. *European Commission Publications Office, Ispra, Italy, EUR report No EUR*, **25229,** 40.
Jung, M.-I., S.-W. Son, H. C. Kim, S.-W. Kim, R. J. Park, and D. Chen, 2019: Contrasting synoptic weather patterns between non-dust high particulate matter events and Asian dust events in Seoul, South Korea. *Atmospheric Environment*, **214,** 116864.
Kim, H. C., and Coauthors, 2016: Synoptic perspectives on pollutant transport patterns observed by satellites over East Asia: Case studies with a conceptual model. *Atmospheric Chemistry and Physics Discussions***,** 1-30.
Kim, G., S. Lee, J. Im, C.-K. Song, J. Kim, and M.-I. Lee, 2021: Aerosol Data Assimilation and Forecast Using Geostationary Ocean Color Imager Aerosol Optical Depth and In-situ Observations during the KORUS-AQ Observing Period. *GIScience and Remote Sensing*, **to be accepted**
Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP global data assimilation system. *Weather and Forecasting*, **24,** 1691-1705.
Lee, H.-H., S.-H. Chen, A. Kumar, H. Zhang, and M. J. Kleeman, 2020: Improvement of aerosol activation/ice nucleation in a source-oriented WRF-Chem model to study a winter Storm in California. *Atmospheric Research*, **235,** 104790.
Lee, H. S., C.-M. Kang, B.-W. Kang, and H.-K. Kim, 1999: Seasonal variations of acidic air pollutants in Seoul, South Korea. *Atmospheric Environment*, **33,** 3143-3152.
Levy, R., S. Mattoo, L. Munchak, L. Remer, A. Sayer, F. Patadia, and N. Hsu, 2013: The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Measurement Techniques*, **6,** 2989-3034.
Li, Z., and Coauthors, 2013: A three-dimensional variational data assimilation system for multiple aerosol species with WRF/Chem and an application to PM 2.5 prediction. *Atmospheric Chemistry and Physics*, **13,** 4265-4278.
Lin, S.-J., and R. B. Rood, 1996: Multidimensional flux-form semi-Lagrangian transport schemes. *Monthly Weather Review*, **124,** 2046-2070.
Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk parameterization of the snow field in a cloud model. *Journal of Applied Meteorology and climatology*, **22,** 1065-1092.
Liu, Z., Q. Liu, H. C. Lin, C. S. Schwartz, Y. H. Lee, and T. Wang, 2011: Three-dimensional variational assimilation of MODIS aerosol optical depth: Implementation and application to a dust storm over East Asia. *Journal of Geophysical Research: Atmospheres*, **116**.
Liu, Z. Q., and F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **128,** 1367-1386.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, **102,** 16663-16682.
Mokhtar, M. M., M. H. Hassim, and R. M. Taib, 2014: Health risk assessment of emissions from a coal-fired power plant using AERMOD modelling. *Process Safety and Environmental Protection*, **92,** 476-485.
Müller, J. F., and G. Brasseur, 1995: IMAGES: A three-dimensional chemical transport model of the global troposphere. *Journal of Geophysical Research: Atmospheres*, **100,** 16445-16490.
Pagowski, M., G. Grell, S. McKeen, S. Peckham, and D. Devenyi, 2010: Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the Weather Research and Forecasting—Chemistry model and Grid-point Statistical Interpolation. *Quarterly Journal of the Royal Meteorological Society*, **136,** 2013-2024.
Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, **120,** 1747-1763.
Peterson, D. A., and Coauthors, 2019: Meteorology influencing springtime air quality, pollution transport, and visibility in Korea. *Elementa: Science of the Anthropocene*, **7**.
Pope Iii, C. A., R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston, 2002: Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*, **287,** 1132-1141.
Randles, C., and Coauthors, 2017: The MERRA-2 aerosol reanalysis, 1980 onward. Part I: System description and data assimilation evaluation. *Journal of climate*, **30,** 6823-6850.
Remer, L. A., and Coauthors, 2005: The MODIS aerosol algorithm, products, and validation. *Journal of the atmospheric sciences*, **62,** 947-973.
Saide, P. E., J. Kim, C. H. Song, M. Choi, Y. Cheng, and G. R. Carmichael, 2014: Assimilation of next generation geostationary aerosol optical depth retrievals to improve air quality simulations. *Geophysical research letters*, **41,** 9188-9196.
Schwartz, C. S., Z. Liu, H. C. Lin, and S. A. McKeen, 2012: Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth. *Journal of Geophysical Research: Atmospheres*, **117**.
Scire, J. S., D. G. Strimaitis, and R. J. Yamartino, 2000: A user's guide for the CALPUFF dispersion model. *Earth Tech, Inc*, **521,** 1-521.
Seangkiatiyuth, K., V. Surapipith, K. Tantrakarnapa, and A. W. Lothongkum, 2011: Application of the AERMOD modeling system for environmental impact assessment of NO2 emissions from a cement complex. *Journal of Environmental Sciences*, **23,** 931-940.
Stauffer, D. R., N. L. Seaman, and F. S. Binkowski, 1991: Use of four-dimensional data assimilation in a limited-area mesoscale model Part II: effects of data assimilation within the planetary boundary layer. *Monthly Weather Review*, **119,** 734-754.
Stowe, L. L., A. M. Ignatov, and R. R. Singh, 1997: Development, validation, and potential enhancements to the second-generation operational aerosol product at the National Environmental Satellite, Data, and Information Service of the National Oceanic and Atmospheric Administration. *Journal of Geophysical Research: Atmospheres*, **102,** 16923-16934.
Talagrand, O., 1997: Assimilation of observations, an introduction (gtspecial issueltdata assimilation in meteology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, **75,** 191-209.
Tie, X., D. Wu, and G. Brasseur, 2009: Lung cancer mortality and exposure to atmospheric aerosol particles in Guangzhou, China. *Atmospheric Environment*, **43,** 2375-2377.
Vestreng, V., G. Myhre, H. Fagerli, S. Reis, and L. Tarrasón, 2007: Twenty-five years of continuous sulphur dioxide emission reduction in Europe. *Atmospheric chemistry and physics*, **7,** 3663-3681.
Weng, F., Y. Han, P. van Delst, Q. Liu, and B. Yan, 2005: JCSDA community radiative transfer model (CRTM). *Proc. 14th Int. ATOVS Study Conf*, 217-222.
Wesely, M., and B. Lesht, 1989: Comparison of RADM dry deposition algorithms with a site-specific

method for inferring dry deposition. *Water, Air, and Soil Pollution*, **44,** 273-293.
Wiedinmyer, C., S. Akagi, R. J. Yokelson, L. Emmons, J. Al-Saadi, J. Orlando, and A. Soja, 2011: The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning. *Geoscientific Model Development*, **4,** 625-641.
Woo, J.-H., and Coauthors, 2020: Development of the CREATE Inventory in Support of Integrated Climate and Air Quality Modeling for Asia. *Sustainability*, **12,** 7930.
Xia, X., J. Min, F. Shen, Y. Wang, D. Xu, C. Yang, and P. Zhang, 2020: Aerosol data assimilation using data from Fengyun-4A, a next-generation geostationary meteorological satellite. *Atmospheric Environment*, **237,** 117695.

Acknowledgments

2009년 UNIST에 입학하고, 2013년에 대학원생이 된 지 8년만인 2021년이 되어서 드디어 박사로서 졸업을 할 수 있게 되었습니다. 지금까지 저를 이끌어주셨던 스승님, 저와 함께 연구를 했던 연구실 선후배들, 그리고 30년이 넘는 시간동안 한 번의 의심도 없이 절 뒤에서 받쳐주시고 응원해 주셨던 부모님께 무한한 감사를 드립니다.

먼저, 2010년부터 인연이 닿아 대학원생이 된 이후 무한한 배려 심으로 저를 이끌어주신 지도교수님이신 이명인 교수님께 감사드립니다. 제가 교수님 속을 이만 저만 썩인 것이 아님을 잘 알고 있습니다. 그럼에도 불구하고 긴 시간 동안 저를 배려해주시고, 격려해 주시며 포기하지 않고 저를 이끌어주신 이명인 교수님께 다시없는 감사의 인사를 드립니다. 교수님이 아니었다면 전 결코 박사가 된다는 것이 어떤 것이며, 또한 어떠한 마음가짐으로 임해야 박사로서 설 수 있을지 결코 알지 못했을 것입니다.

그리고 부족하지만 최선을 다 하여 준비한 박사논문을 심사하고 평가해주신 심사위원분들께도 이 기회를 빌려 감사의 인사를 드립니다. 이명인 교수님 못 지 않게 오래 뵈었던 최성득 교수님, 국립환경과학원에서 시작되어 현재 UNIST에서 함께 하고 계신 송창근 교수님, 그리고 이명인 교수님과의 인연으로 심사를 함께 해 주셨던 부산대학교의 김철희 교수님과 서울대학교의 박록진 교수님께 감사드립니다.

아울러, 2011년 연구실 인턴으로 시작하여 현재까지 저와 연구실 생활을 함께했던 선후배들에게도 감사드립니다. 연구실 인턴 시절부터 시작하면 무려 10년을 한 연구실에 있었는데, 그 동안 만나서 함께하고, 먼저 나가고 새로 들어오는 사람들 모두와 소중한 인연을 맺게 되었습니다. 지도교수님 다음으로 제게 가장 많은 가르침을 주셨던, 지금은 플로리다에 계신 김동민 박사님, 그리고 여러 가지 세심한 것들까지 많이 알려주셨던 김혜림 박사님, 특유의 안정적인 분위기로 연구실 분위기를 잘 환기시켜 주셨던 김성윤 형, 동기 중에서 가장 빠르게 스스로의 연구 주제를 잡고 가장 먼저 졸업할 수 있는 추진력을 가진 강대현 박사, 그리고 스스로가 정한 연구 주제에 대한 망설임이 없이 기반에서부터 시스 템을 갈고 닦아 졸업하였던 서은교 박사, 랩실의 또 다른 축인 머신러닝을 가장 먼저 담당했던 민상이와 귀여움을 독차지한 선주, 저를 가장 많이 도와주었으며 저 또한 많은 도움을 주었던 혁재형과 흔한 인연은 아니었으나 역시 도움을 주면서 함께 모델을 공부 하였었던 Son, 그리고 박사후 연구원으로서 잠시 오셔서 연구실 사람들 모두에게 도움을 주셨던 박명숙 박사님까지, 저보다 먼저 졸업하여 나가신 분들과 부산대

에서 박사후 연 구원으로 오셔서 현재 연구교수까지 되시고 많은 도움을 주신 이준리 박사님, 먼저 졸업했지만 이후 자리가 정해지기까지 학교에 남아서 학생들에게 도움을 주고 있는 최낙빈 박사, 부사수로서, 그리고 지금은 스스로 박사가 되기 위해 많은 노력을 기울이고 있는 승희, 현재 랩장으로서 랩의 일을 도맡아 하는 지해와 폭염 및 여타 연구를 중점적으로 하는 선래, Son 이후에 또다른 인연이 닿아 만나게 된 Naser, 그리고 현재 인턴으로 들어와 스스로의 꿈을 키우고 있는 내현이, 병헌이, 혜린이, 승석이, 건희, 그리고 하경이와 같은 아직 연구실에 남아있는 사람들까지 정말로 많은 인연을 만들 게 되었고, 서로 도와가면서 연구실 생활을 지속할 수 있었습니다. 이 수많은 인연에 감사드리며, 이후에도 긍정적인 인연이 계속 유지될 수 있었으면 좋겠습니다.

마지막으로 저를 끝까지 믿고 아껴주시고 지지해주셨던 부모님께 정말 큰 감사를 드립니다. 박사학위에 도전한다는 것 자체가 결코 쉬운 일이 아닐뿐더러, 제가 동기들보다 시일이 많이 지체된 후에 졸업을 하는 것인 만큼 부모님께서도 더욱 애타는 심정으로 저를 기다리셨을 것이라고 생각됩니다. 앞에서는 지도교수님, 옆에서는 연구실 선후배들이 함께했다면 뒤에서 언제나 응원을 보내주셨던 분들은 바로 부모님이셨습니다. 그에 힘입어 8년의 연구실 생활을 이렇게 긍정적으로 마칠 수 있어서 정말로 다행이라고 생각합니다.

그리고 미처 소개하지 못하였지만, 여러 연구를 함께 하며 서로 도움을 주고받았던 타 연구실, 대학, 기관의 인연들에게도 감사의 말씀을 드립니다.

단순히 지구과학에 꿈이 있었을 뿐인 소년에서 시작하여 이후 박사로서 스스로 발을 딛고 서기까지 정말로 많은 우여곡절이 있었고 심지어 졸업연장과 휴학까지 하면서도 포기하지 않고 최선을 다하여 유종의 미를 거두었습니다. 그리고 이제 비로소 세상에 나갈 준비가 되었다고 생각합니다. 다시 한번, 지도교수님이신 이 명인 교수님, 그리고 UNIST Climate Environmental Modeling Laboratory, UCEM의 모든 선후배, 동료들과 부모님께 감사드립니다.

2021/06/15

김 강한 드림